Où va l’intelligence artificielle ?

Bertrand Thirion

bertrand.thirion@inria.fr

Pourquoi parle-t-on autant d'IA aujourd'hui ?

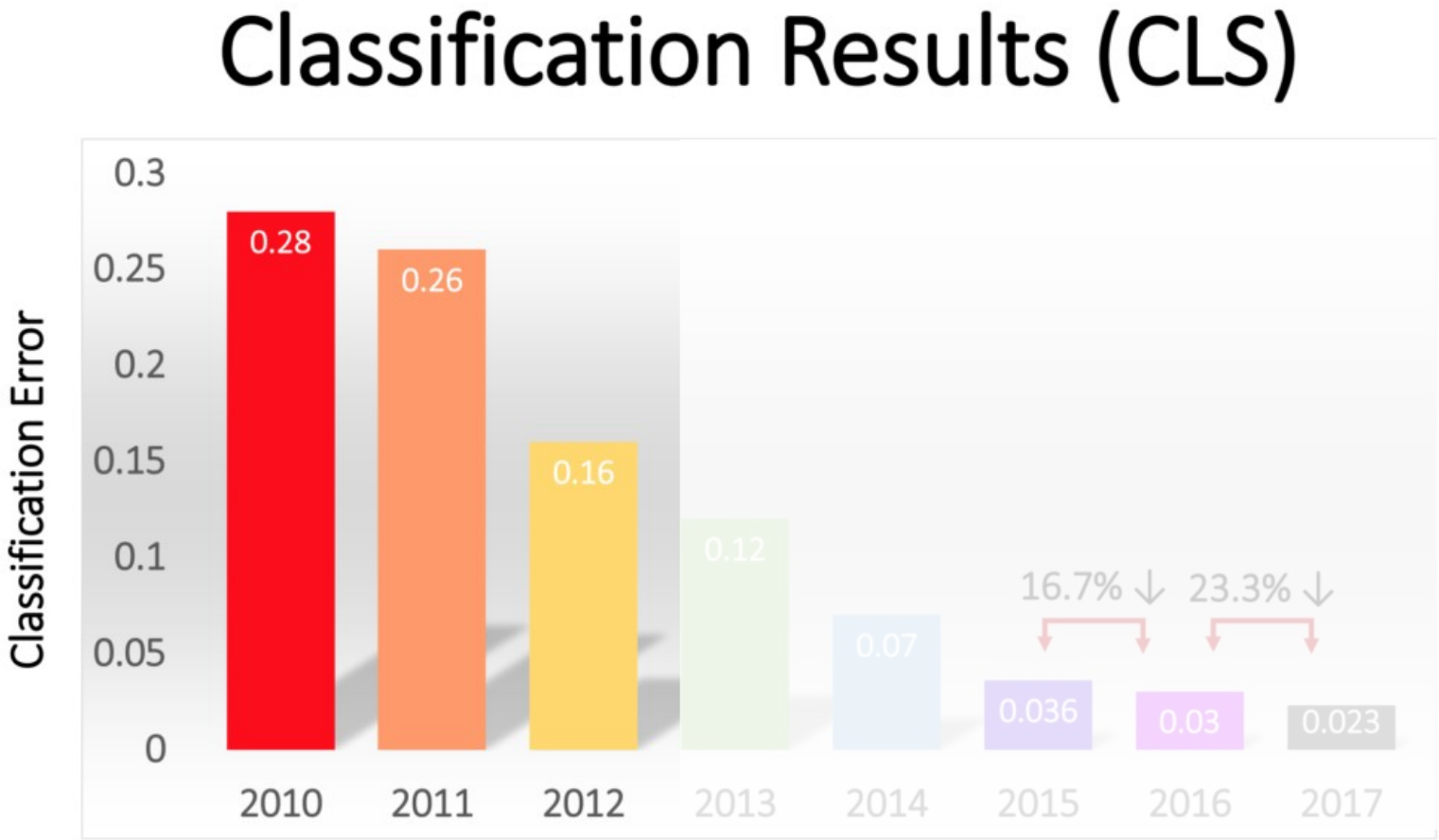
- Lame de fond (90's -aujourd'hui): La "société des données"
  - Internet
  - Numérisation généralisée des données
  - Partage des données
  - Industrialisation du secteur de la donnée (GAFAM...)
- Singularité: ILSVRC-2012

ILSVRC-2012 : imagenet

ILSVRC-2012

- Compétition de classification d'image: identifier la sémantique d'un image, sur un grand jeu de données (Imagenet)

Brusque saut de performance

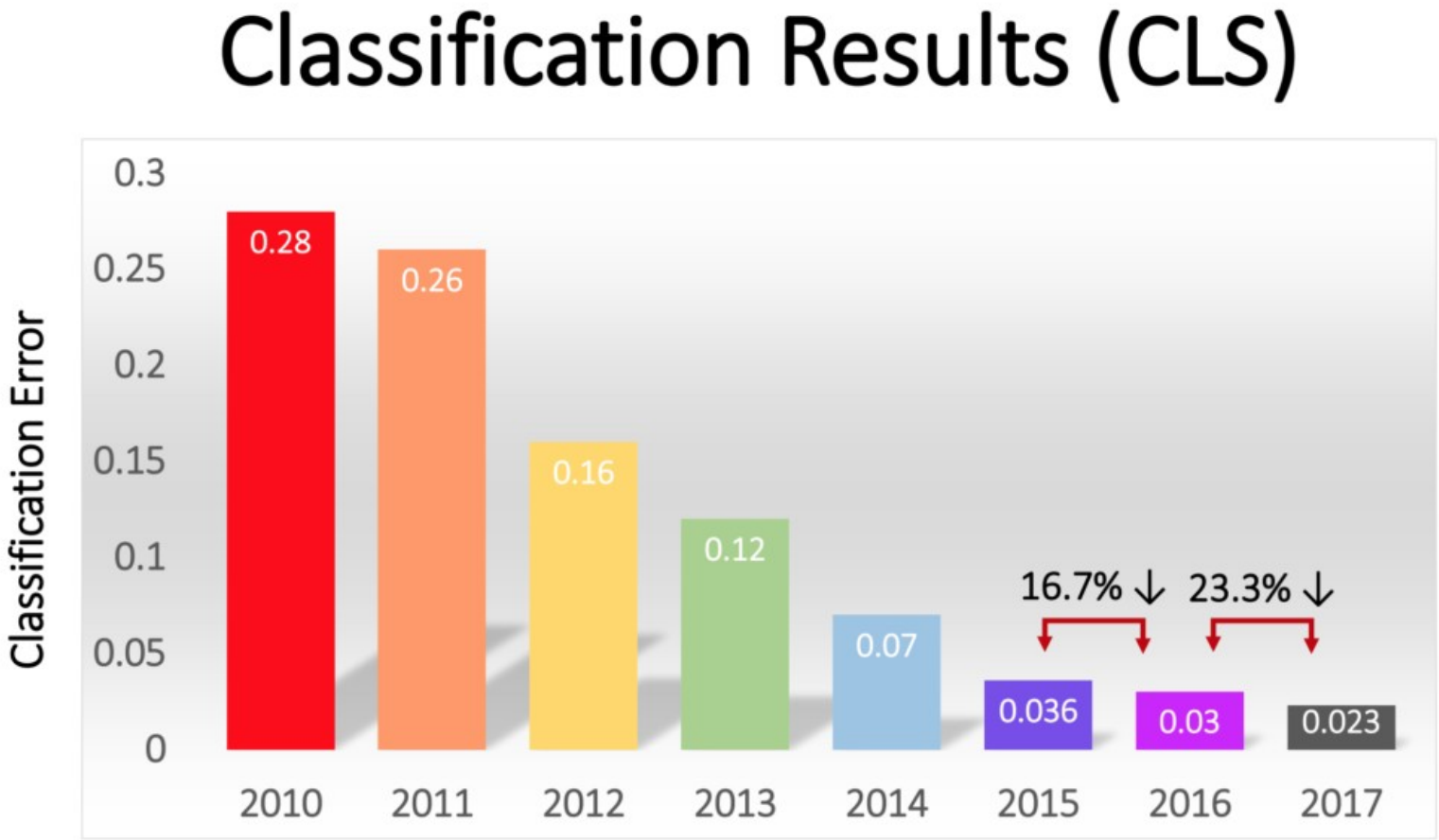


https://www.bulentsiyah.com/imagenet-winning-cnn-architectures-ilsvrc

ILSVRC-20**

- Compétition de classification d'image: identifier la sémantique d'un image, sur un grand jeu de données (Imagenet)

Changement qualitatif



L’essor d’une nouvelle technologie

- Turing award 2018 accordé à Bengio, LeCun et Hinton

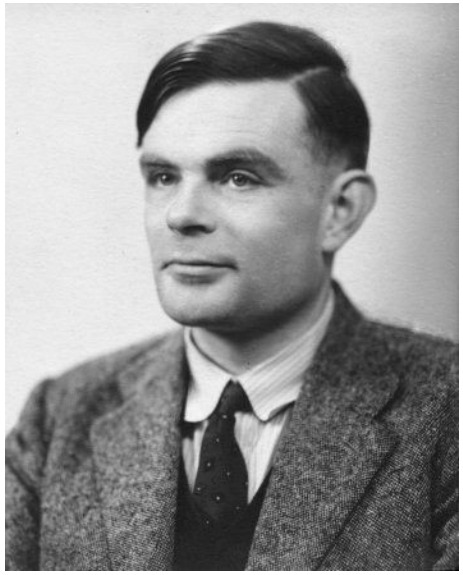
- L’apprentissage profond (deep learning)

Plan

- Définition & histoire de l’intelligence artificielle
- Apprentissage automatique et l’intelligence artificelle
- Faut-il s’inquiéter de l’omniprésence des systèmes d’IA ?

“Intelligence artificielle”

“L'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de **simuler** l'intelligence humaine” Larousse



Brève histoire de l'IA

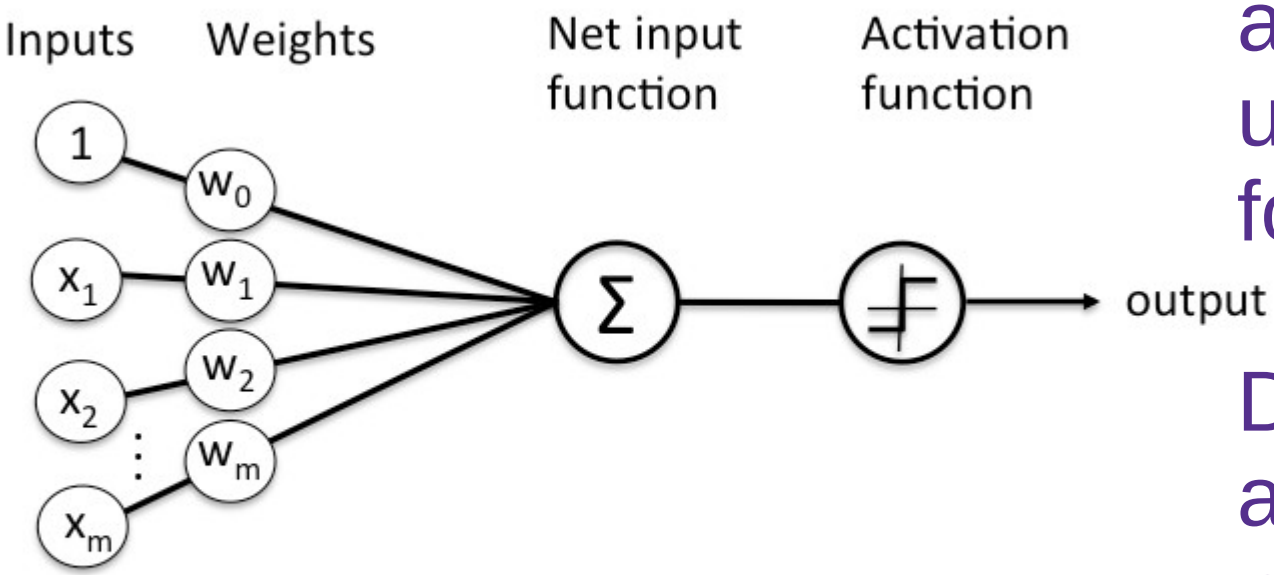
- Alan Turing, 1950: une machine peut-elle penser ? à partir de quand devient-elle « consciente » ?

1956: Conférence de Dartmouth, 20 chercheurs,

- permières solutions algorithmiques
- terme "Intelligence artificielle"

Le Perceptron

- Franck Rosenblatt, 1958



Schematic of Rosenblatt's perceptron.

Modèle **bio-inspiré,** approximateur universel de fonctions

Devient puissant avec une structure **multi-couche** (*multi-layer perceptron*)

L'IA et ses hivers

https://www.afas.fr/vers-un-deploiement-industriel-de-lintelligence-artificielle

2 hivers: 1973-1980, 1987-1993, puis IA en retrait dans les années 2000

L’âge des succès: Kasparov 1997

- Victoire aux échecs de Deep Blue (IBM) contre Kasparov
- Pas de nouveaux paradigme
- Application minutieuse des techniques d'ingénierie et puissance de calcul

L’âge des succès: Jeopardy, 2011

- Progrès en traitement de la langues
- Connexion avec les bases de connaissance
- Puissance de calcul

L’âge des succès: Alphago, 2017

- Système de deepMind bat le champion du monde au Go
- Entrainé en jouant contre lui même
- Combine apprentissage et parcours de graphe
- Gagne aussi aux échecs etc.

Pourquoi l'IA réussit son pari ?

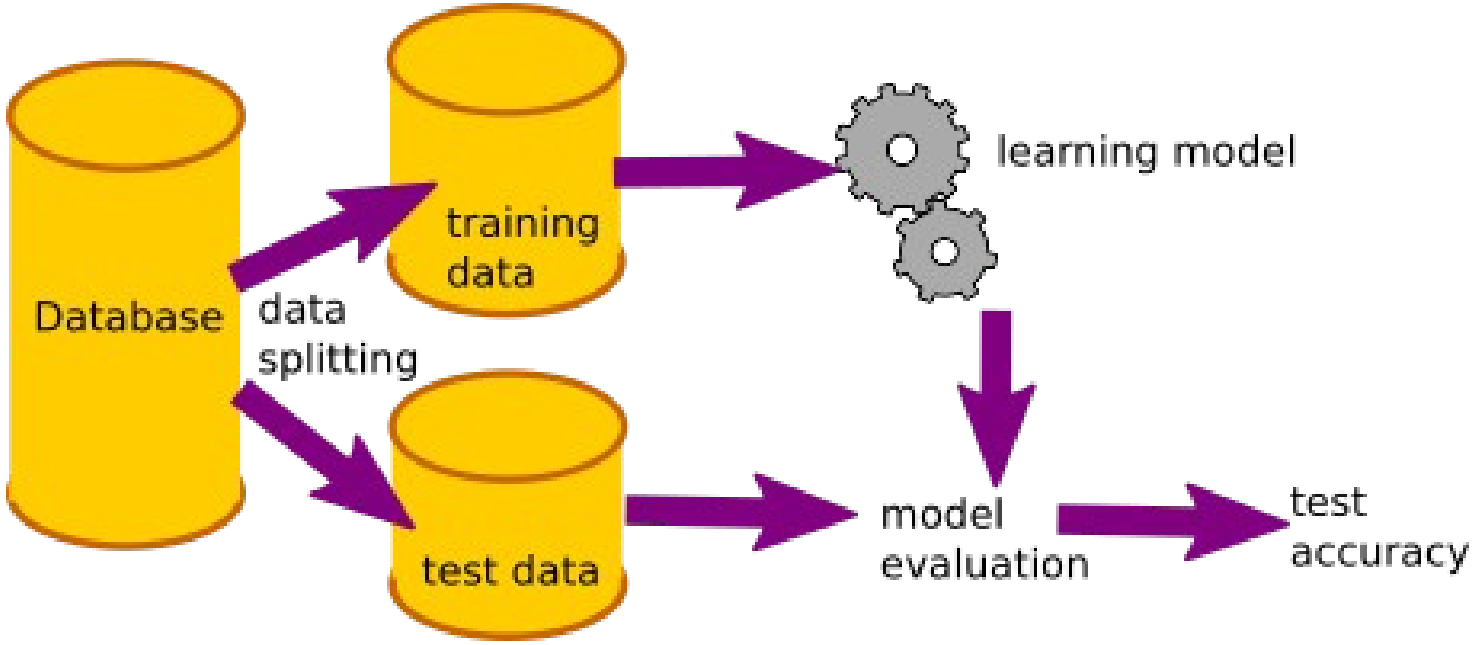
- Augmentation de la puissance de calcul
- Mise au point de jeux de données gigantesques
  - Imagenet en 2009
- Abandon des méthodes formelles pour des méthodes statistiques
- La culture du *benchmark*

Plan

- Définition & histoire de l’intelligence artificielle
- Apprentissage automatique et l’intelligence artificelle
- Faut-il s’inquiéter de l’omniprésence des systèmes d’IA ?

Le bras armé de l'IA: l'apprentissage automatique

- Apprentissage automatique = statistiques sur des données complexes
- Modèles à entrainer, puis valider

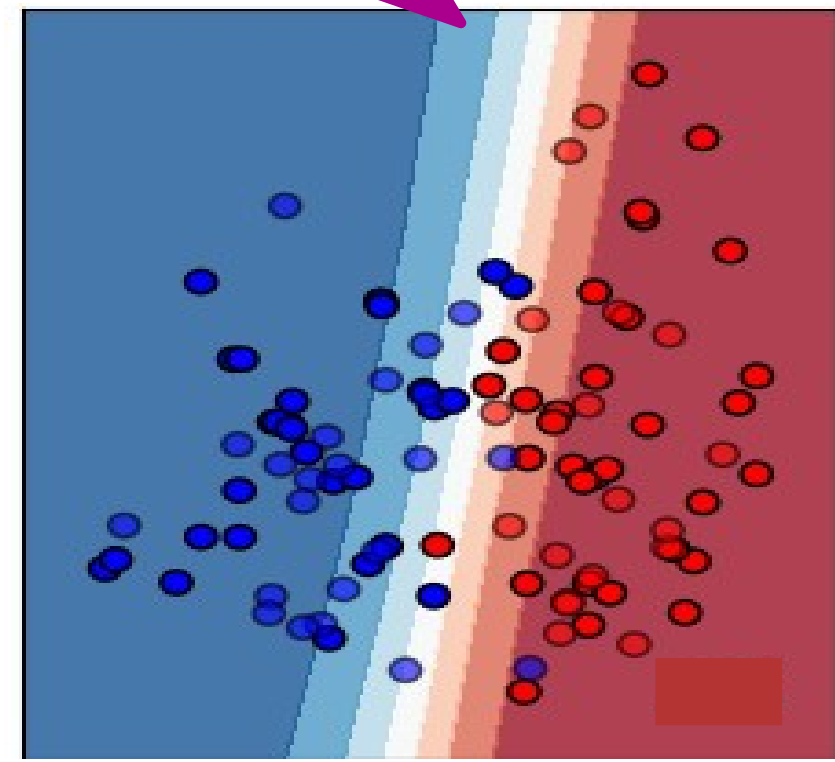


L’apprentissage automatique

L’apprentissage supervisé

Fonction F(x)

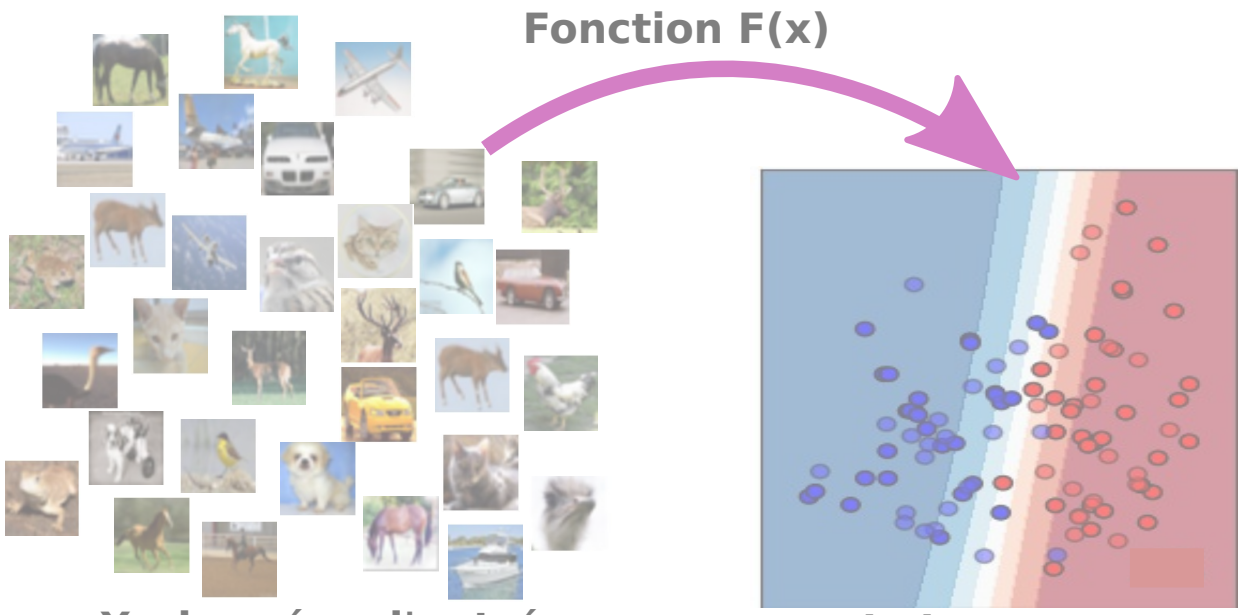
X: données d'entrée



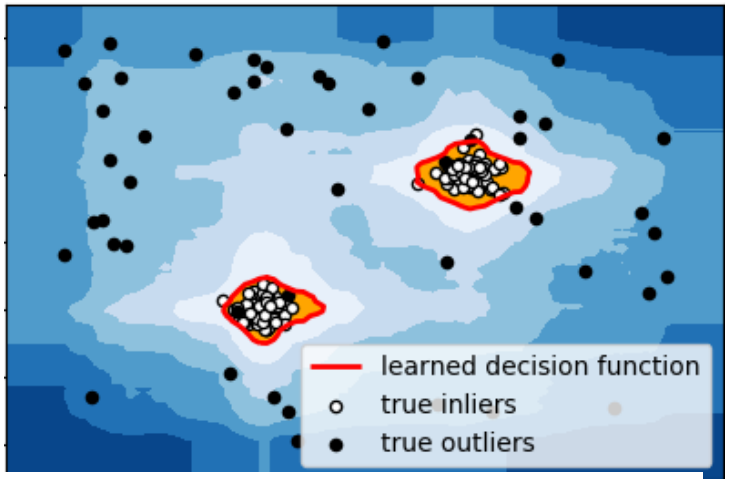
Y: Labels

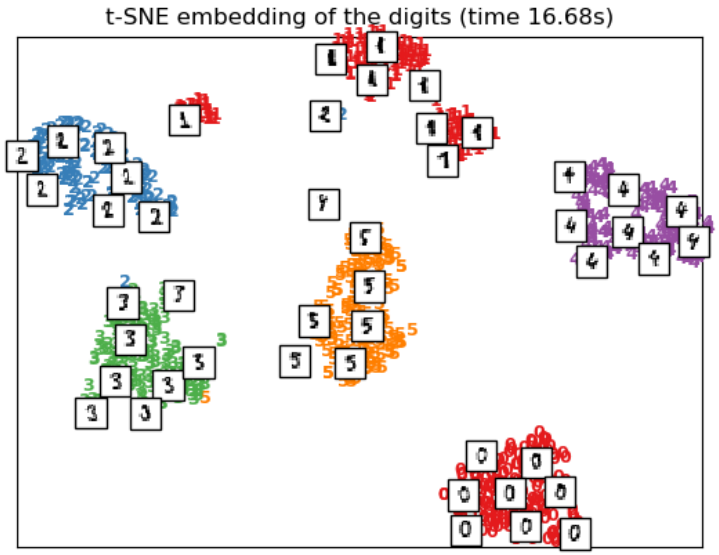
L'apprentissage automatique

Apprentissage supervisé



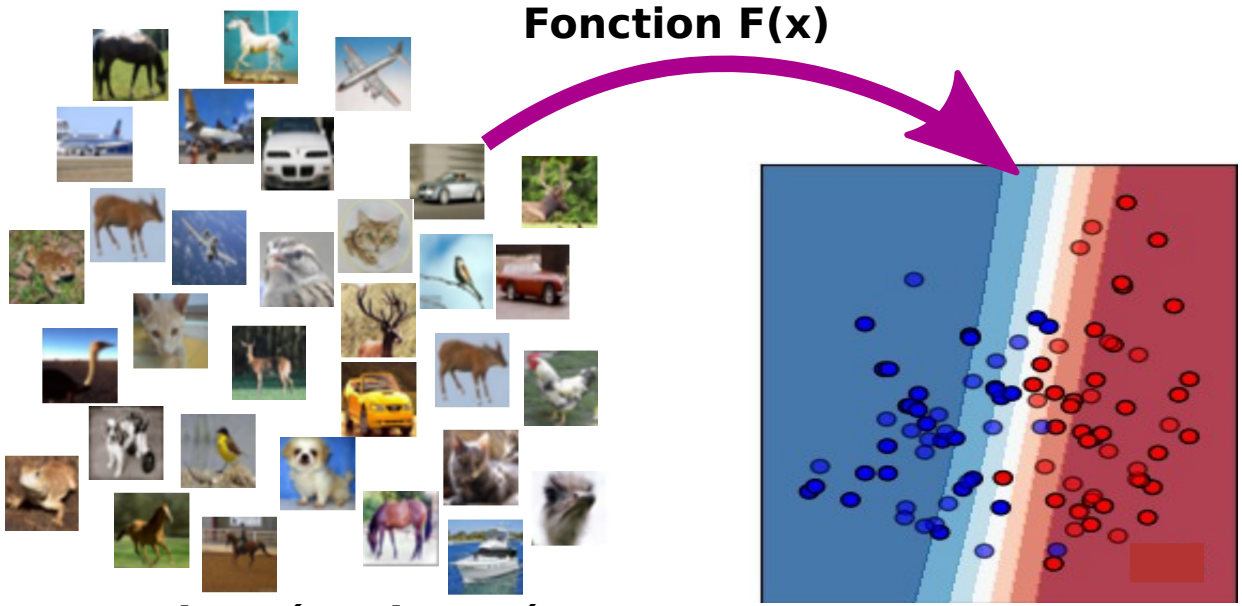
Apprentissage non-supervisé



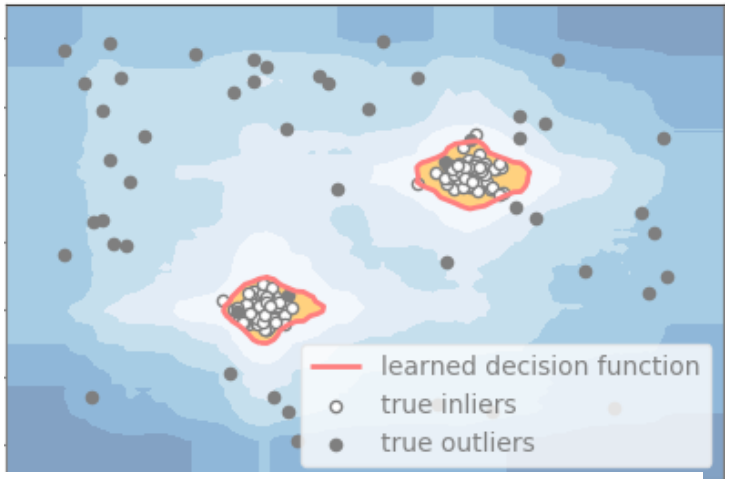


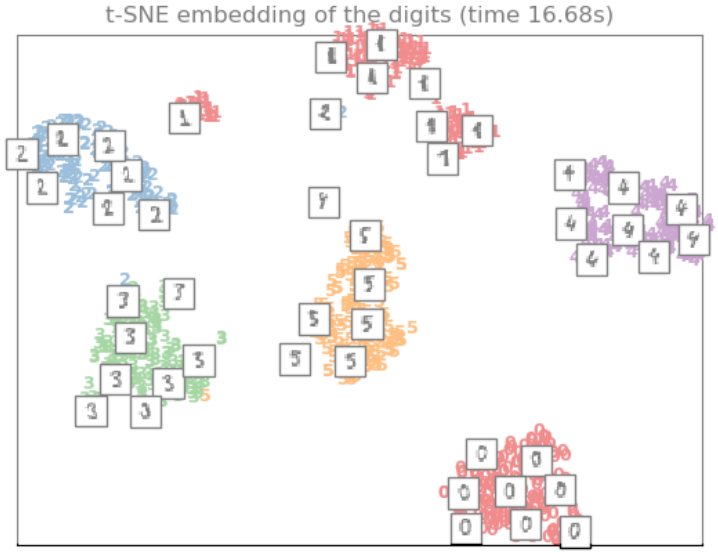
A quoi l'apprentissage est-il bon ?

✔ Apprentissage supervisé



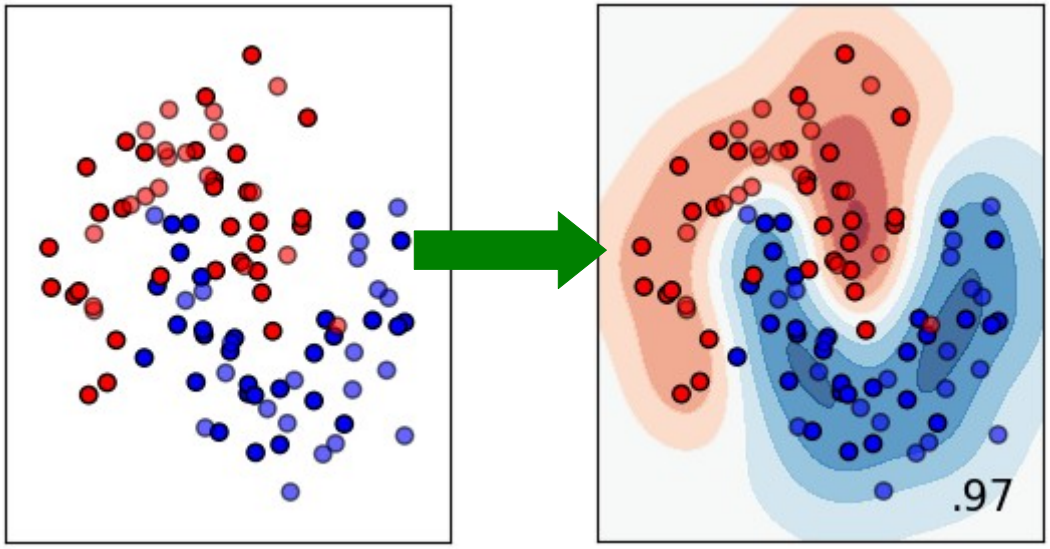
✗ Aprentissage non-supervisé





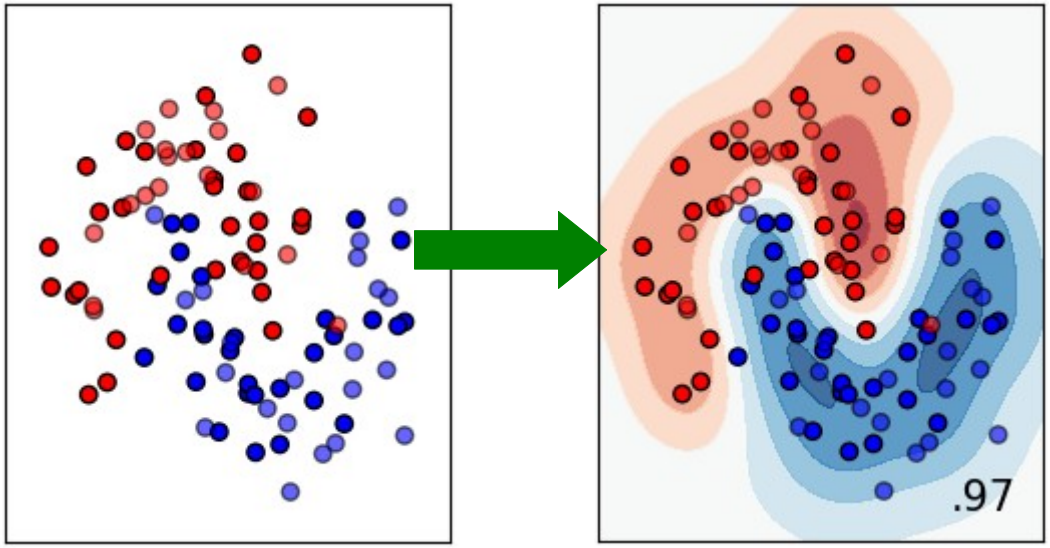
Comment marche l'apprentissage ?

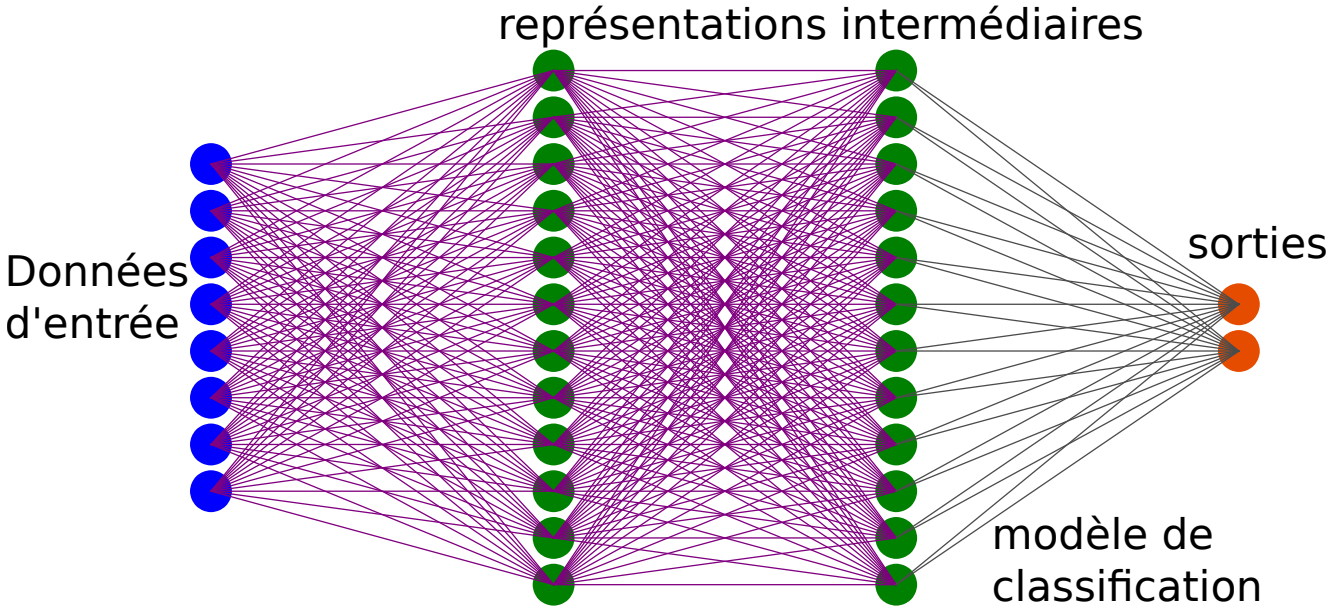
Apprentissage supervisé



A quoi l'apprentissage est-il bon ?

Apprentissage supervisé



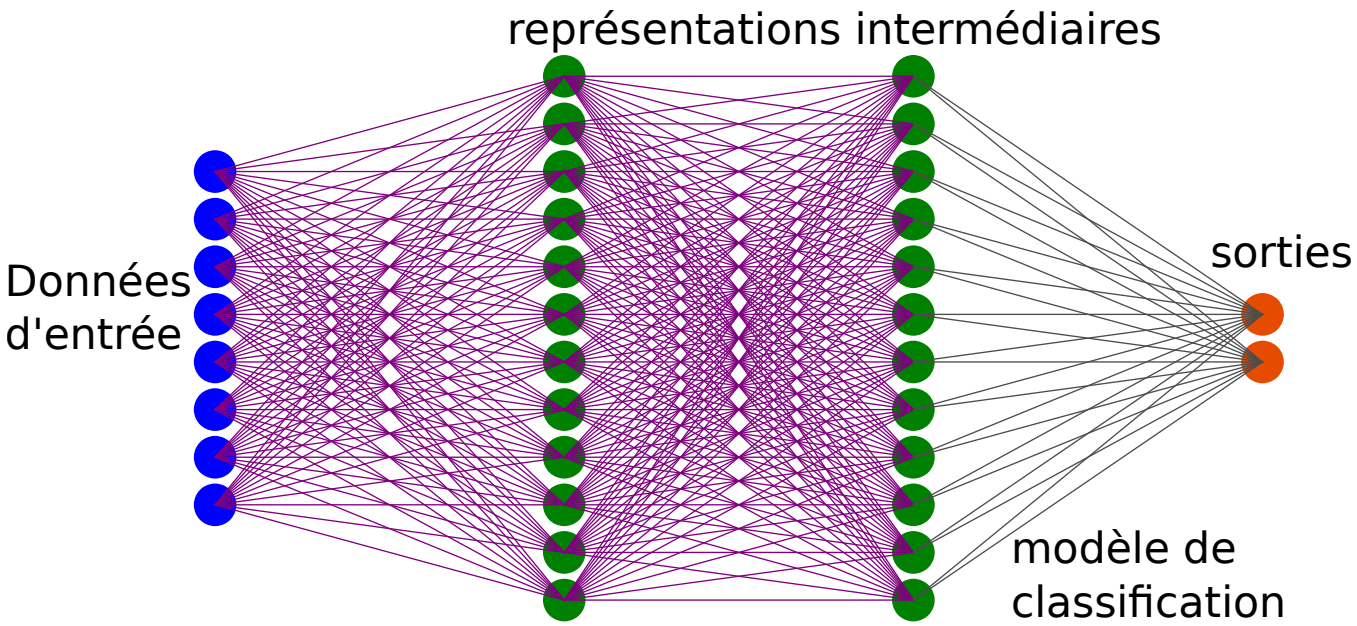


2 étapes

- créer de bonnes représentations des données
- classifier à partir des représentations

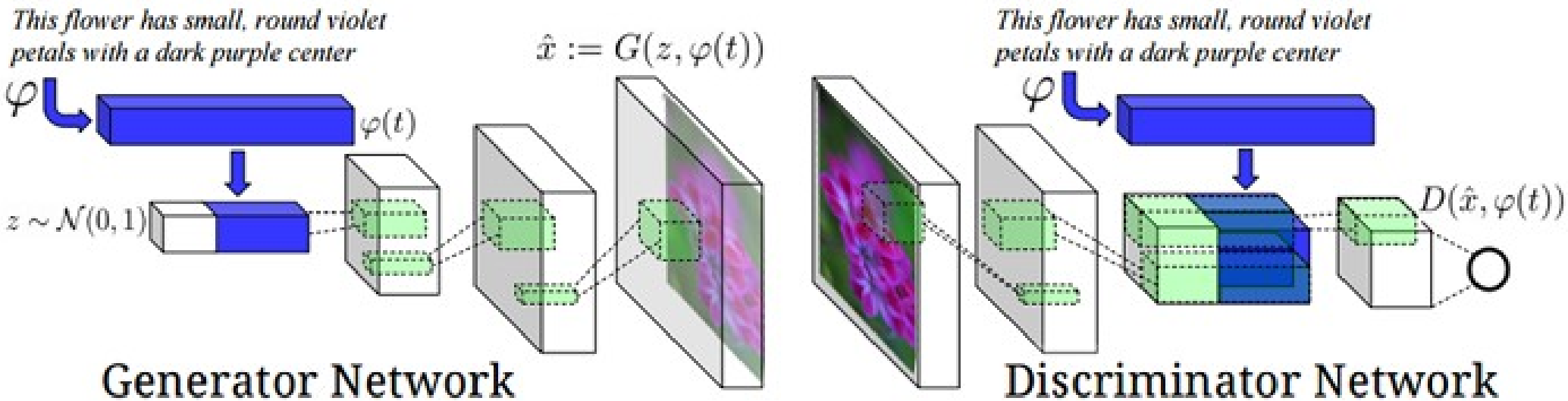
L’apprentissage de bout en bout

Fait manuellement jusqu’à récemment:
Modélisation = traitement de données sophistiqué



Fait désormais par le réseau de neurones seul, par descente de gradient

Les réseaux génératifs adversariaux



[Goodfellow et al. 2014]

https://www.kdnuggets.com/2016/10/deep-learning-research-review-generative-adversarial-networks.html

Les réseaux génératifs adversariaux

https://www.whichfaceisreal.com/methods.html

Le traitement du language

- Réussite tardive (2017)
- Architecture des **Transformeurs** (GPT3, Bert)
- Entrainement efficaces sur de très gros corpus
- Traduction, génération, complétion de texte

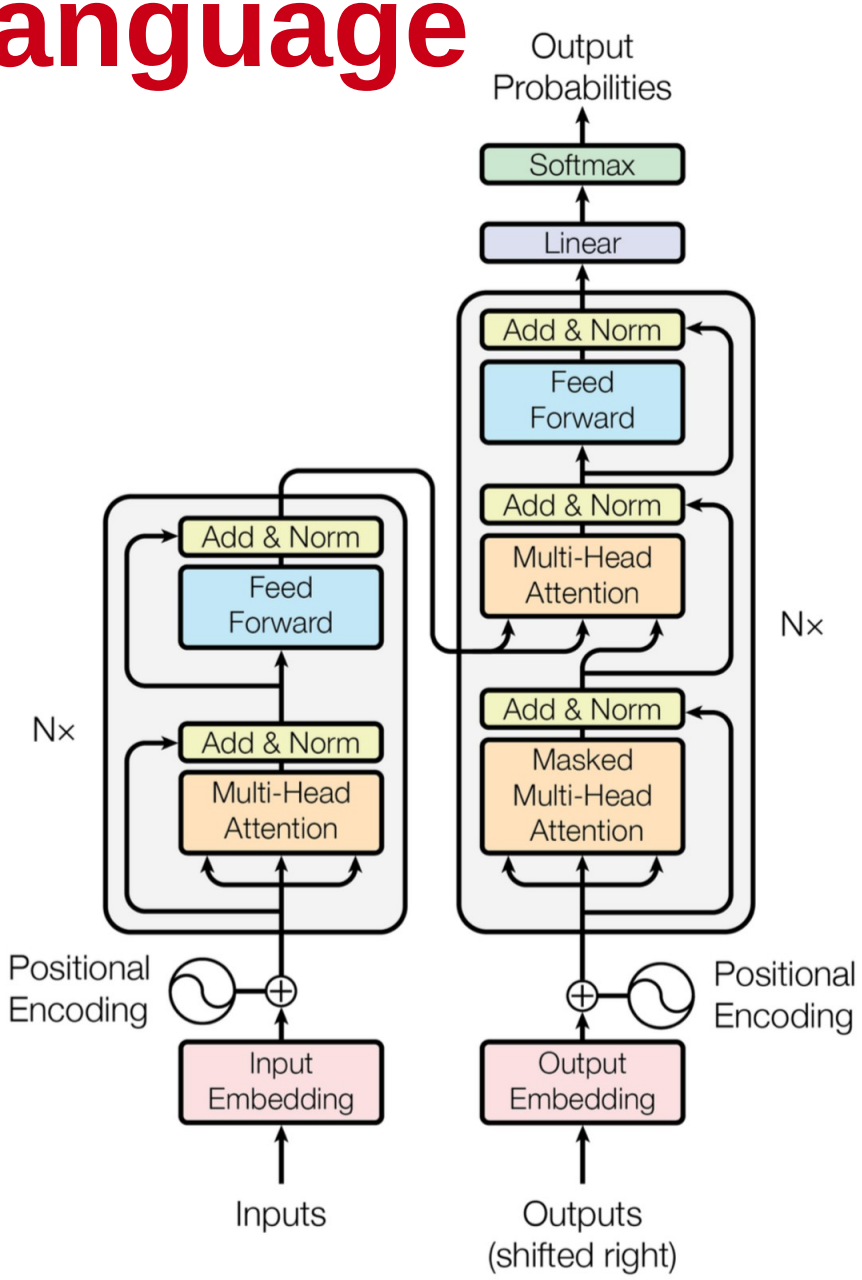


Figure 1: The Transformer - model architecture.

Des modèles de langage de plus en plus puissants: PALM (Google)

Explaining a joke

Prompt

Explain this joke:

Joke: Did you see that Google just hired an eloquent whale for their TPU team? It showed them how to communicate between two different pods!

Model Response

Prediction: TPUs are a type of computer chip that Google uses for deep learning. A "pod" is a group of TPUs. A "pod" is also a group of whales. The joke is that the whale is able to communicate between two groups of whales, but the speaker is pretending that the whale is able to communicate between two groups of TPUs.

L'IA "artistique"

DALL·E 2 can create original, realistic images and art from a text description. It can combine concepts, attributes, and styles.

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

riding a horse lounging in a tropical resort in space playing basketball with cats in space

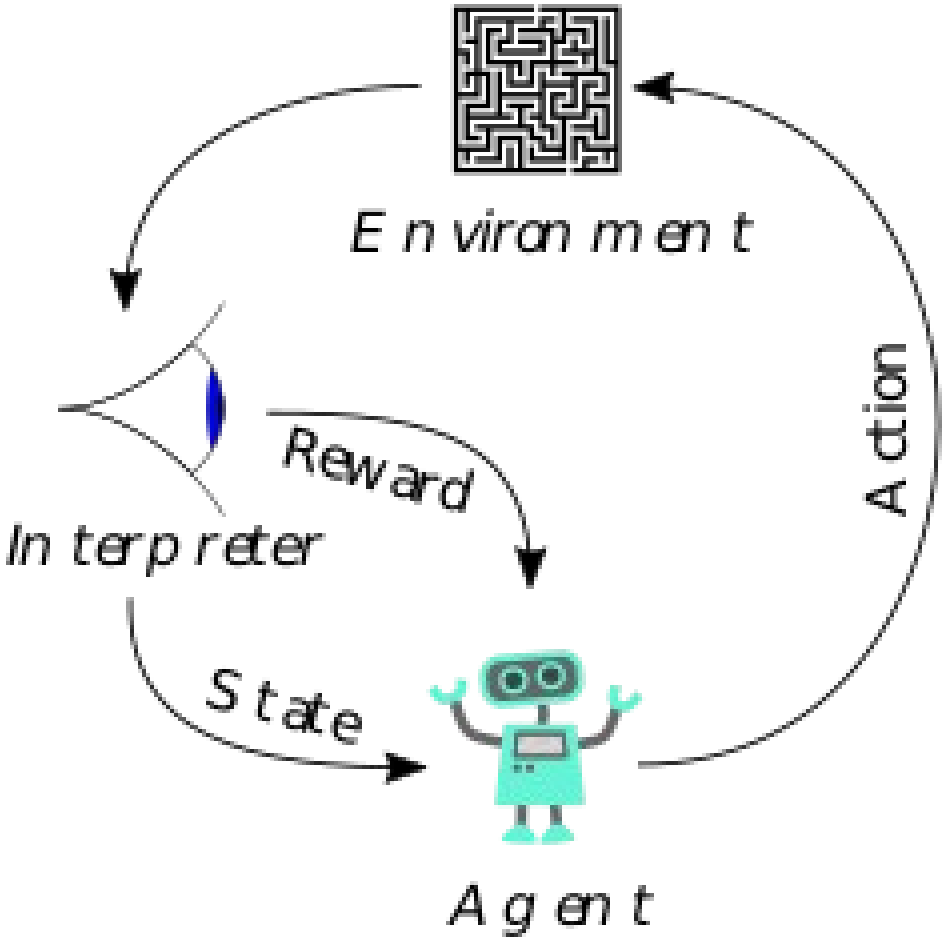
in a photorealistic style in the style of Andy Warhol as a pencil drawing

→

DALL·E 2

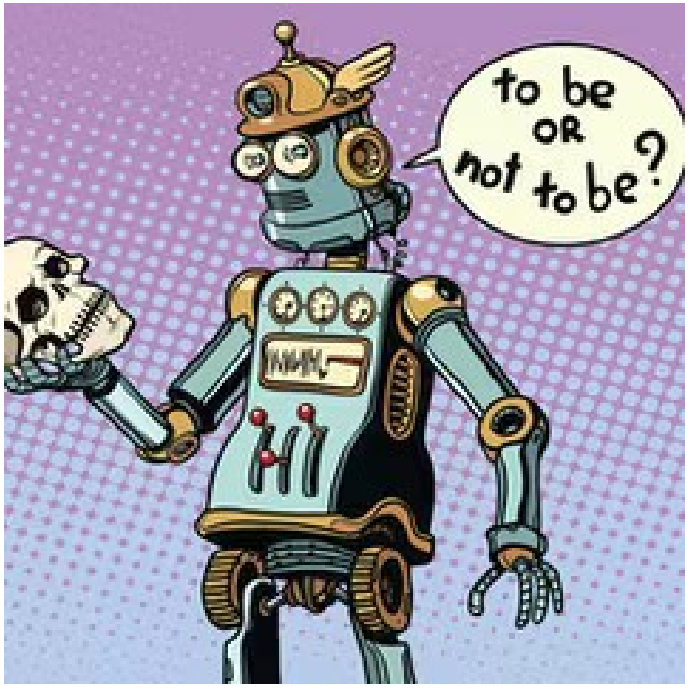
L’apprentissage par renforcement

- Un agent évoluant dans un environnement changeant et recevant des récompenses, apprend à adapter sa conduite par essai et erreur
- Dilemme exploration/exploitation



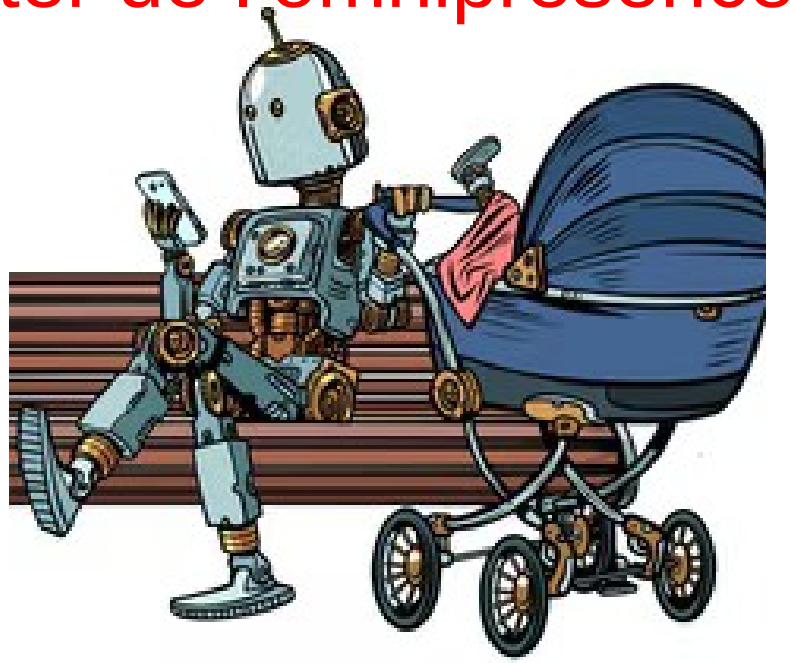
De l'IA faible à l'IA forte

- IA faible: spécialisation pour des tâches simples, stéréotypées
- IA forte:
  - capacité à faire de multiples tâches
  - Adaptation à des environnements évolutifs
  - Prise en compte des autres acteurs
  - Capacité à apprendre continument,
  - adapte ses objectifs



Plan

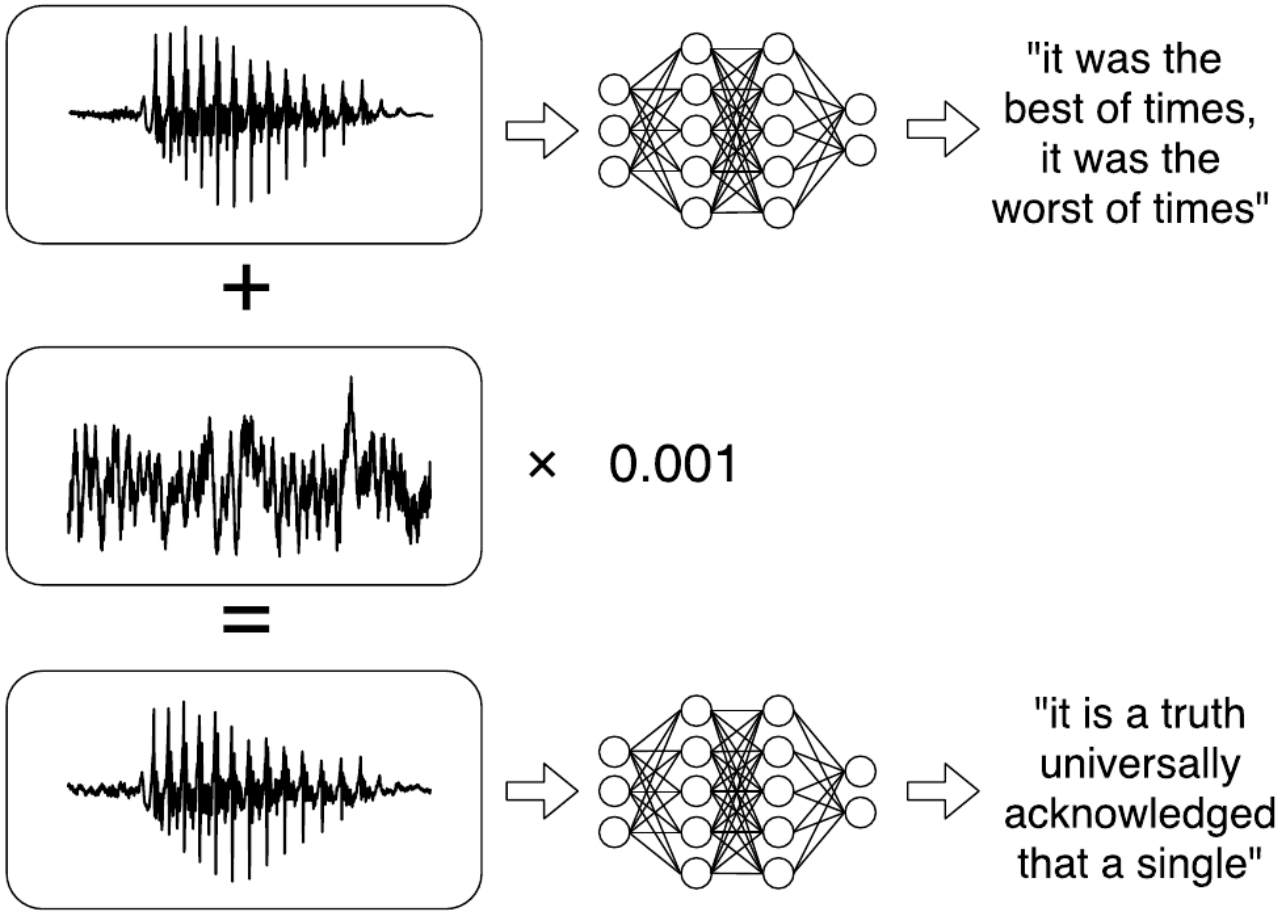
- Définition & histoire de l’intelligence artificielle
- Apprentissage automatique et l’intelligence artificelle
- Faut-il s’inquiéter de l’omniprésence des systèmes d’IA ?

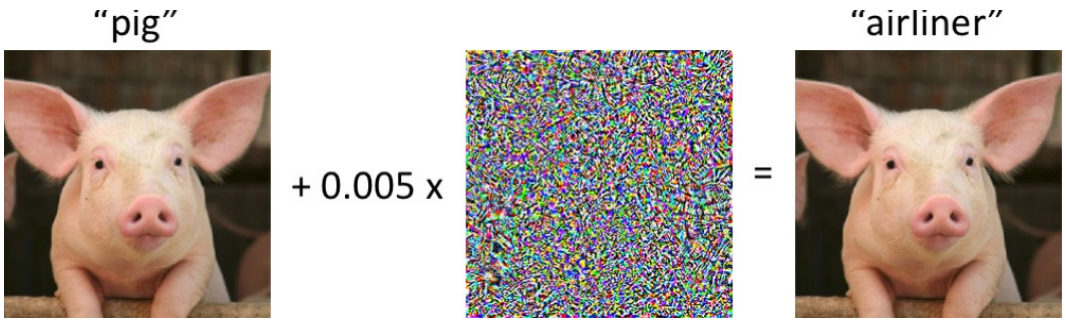


Faut-il s'inquiéter de l'omniprésence de l'IA ?

- L'IA = système expert: quelle contre-expertise ?
- Des agents intelligents qui nous servent … ou nous manipulent
- Robustesse/adaptabilité limitée
- Les biais: l'IA ne crée par les problèmes, mais elle peut les amplifier/perpétuer
- effet d'autorité / sur-confiance: on se repose excessivement sur les sorties des IAs

Erreurs aberrantes





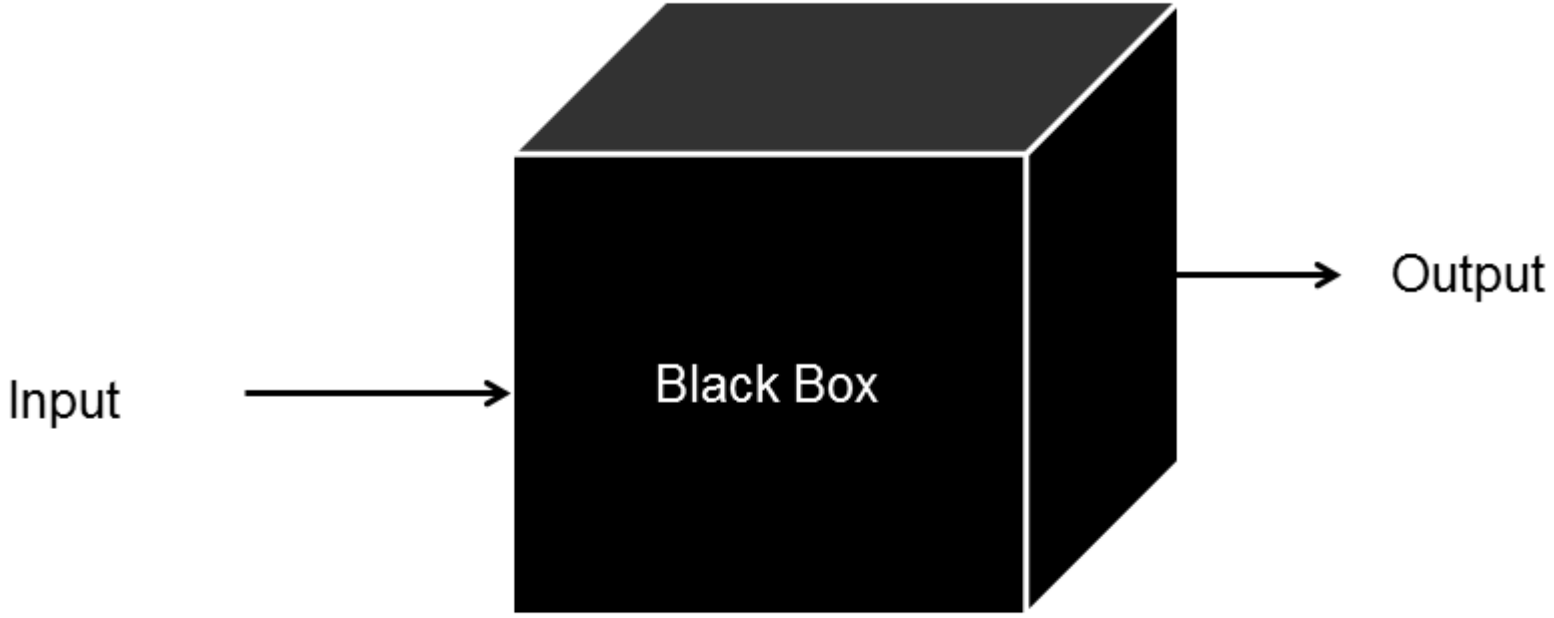
Echecs rentissants

- Projet "Watson for Oncology" d'IBM (62M$) annulé après et des recommandations de traitements dangereux
- Le Chatbot IA de Microsoft corrompu par des trolls sur Twitter
- L'identification faciale d'Apple trompée par un masque 3D
- Amazon arrête son IA pour le recrutement parce que ses ingénieurs l'ont entraînée à être misogyne
- Le logiciel de reconnaissance faciale d'Amazon associe 28 membres du Congrès américain à des criminels.

https://www.lexalytics.com/lexablog/stories-ai-failure-avoid-ai-fails-2020

L'IA "boîte noire"

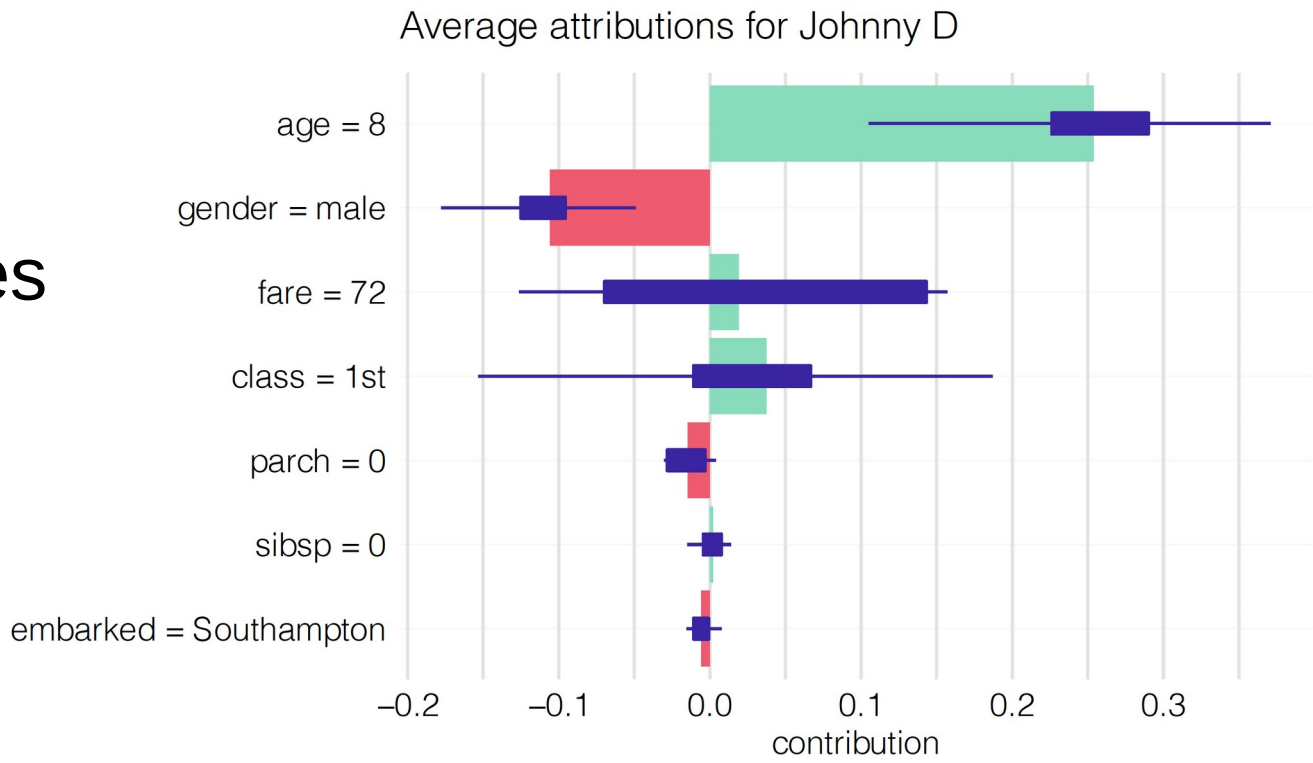
- Modèles d'IA extrêmement complexes: danger de ne pas (chercher à) comprendre la décision



Internal behavior of the code is unknown

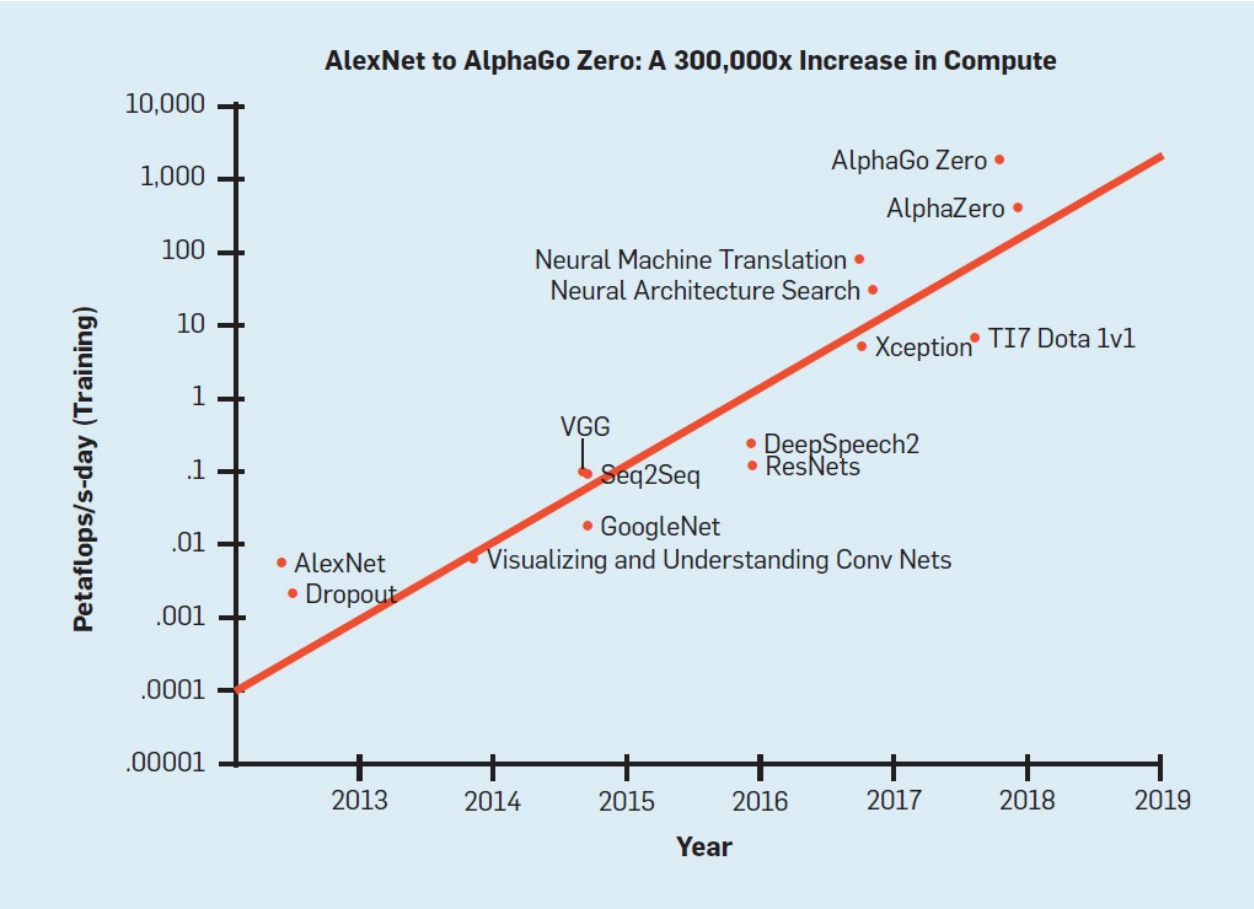
Ouvrir la boite noire

- IA explicable:
  - Identifier les dimensions prédictives dans les données
  - Identifier pourquoi cette instance en particulier est classifiée comme elle l'est



Exemple: titanic dataset
[Molnar, interpretable machine learning]

Coût énergétique/écologique de l'IA



Quantité de données et #paramètres astronomiques

- Désastre écologique! 300T CO2 pour entrainer un modèle !
  + Résultats non-reproductibles

La logique des big data: la valeur dans les données

- Coût majeur pour IA performantes: les données (qualité des annotations)
- Apparition d'un prolétariat de la donnée (mechanical turk etc.)
- "Quand tout est gratuit, c'est vous le produit"
- Conséquences:
  - concentration, la domination de quelques acteurs
  - Nudging des usagers par des produits *gratuits* (gmail, google map)

La culture du logiciel libre

https://scikit-learn.org/stable/

scikit learn Install User Guide API Examples More

scikit-learn

Machine Learning in Python

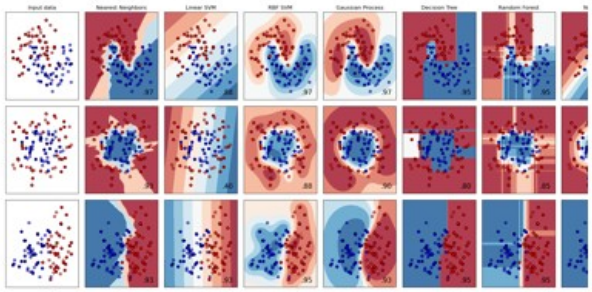
Getting Started | Release Highlights for 1.0 | GitHub

- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.
Algorithms: SVM, nearest neighbors, random forest, and more...

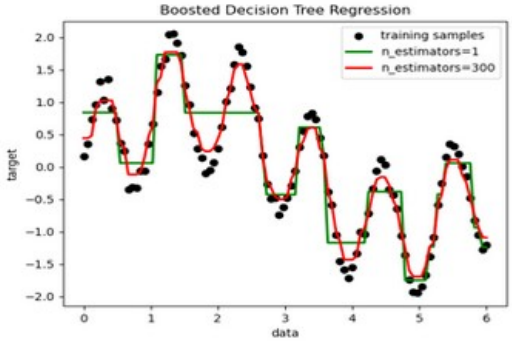


Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, nearest neighbors, random forest, and more...

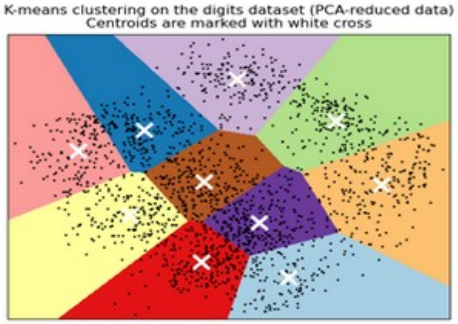


Examples

Clustering

Automatic grouping of similar objects into sets.

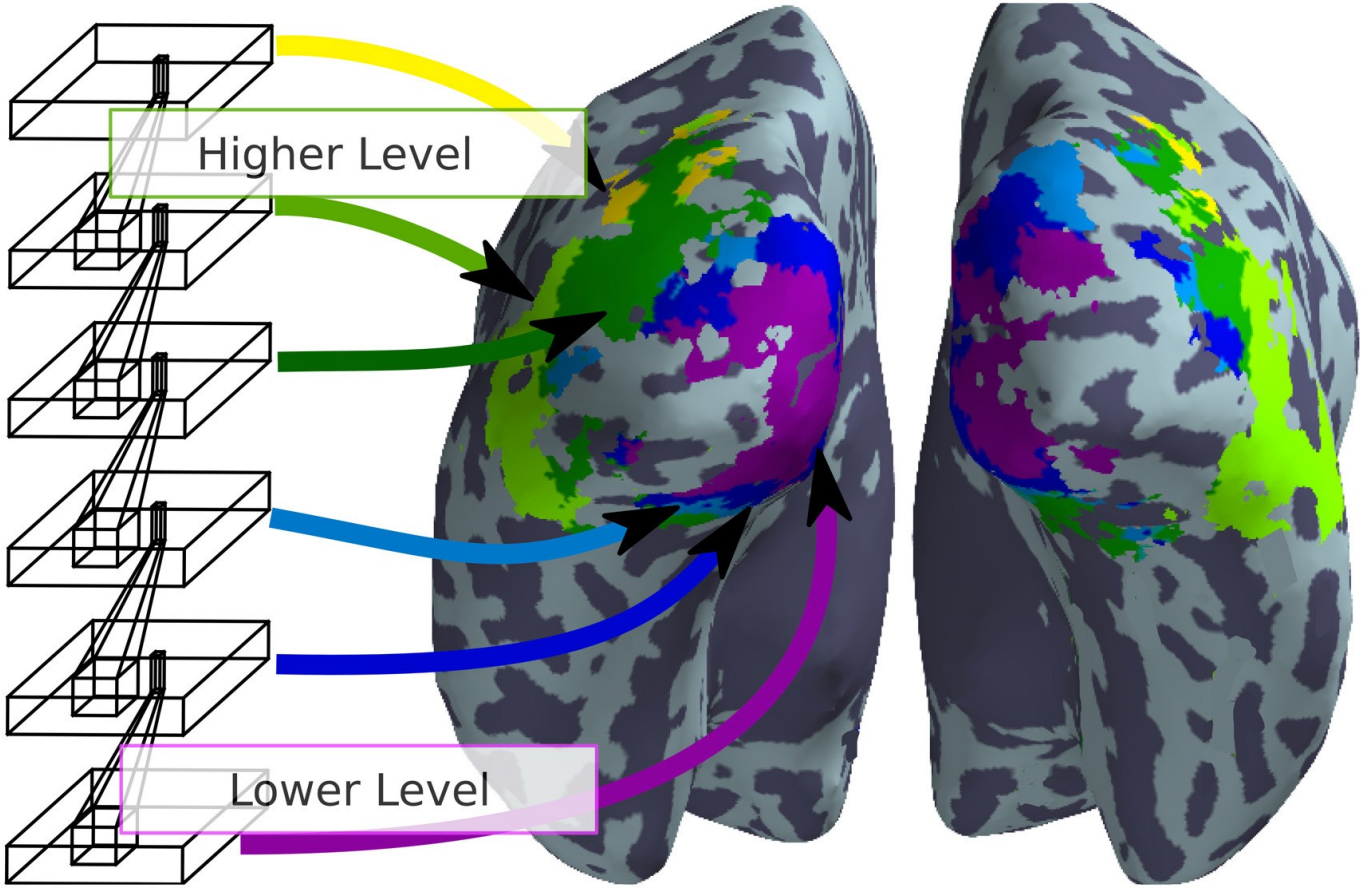
Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, and more...



Examples

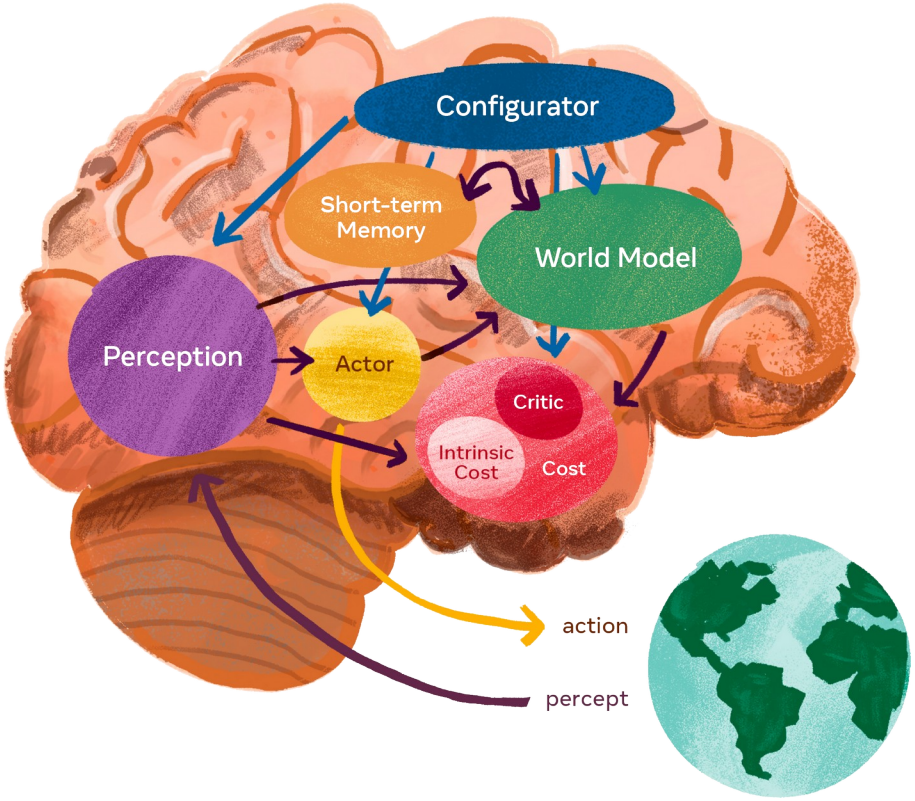
L’IA et le cerveau

- L’IA comme modèle de grandes fonctions cognitives: la reconnaissance d’objet, le traitement du langage



L'IA et le cerveau

- Vers un IA forte ("autonome", LeCun): mise au point de système qui ont un modèle du monde, une capacité de planification, des objectifs internes qui peuvent évoluer
- Inspiration directe du cerveau



https://ai.facebook.com/blog/yann-lecun-advances-in-ai-research/

Education à l'IA

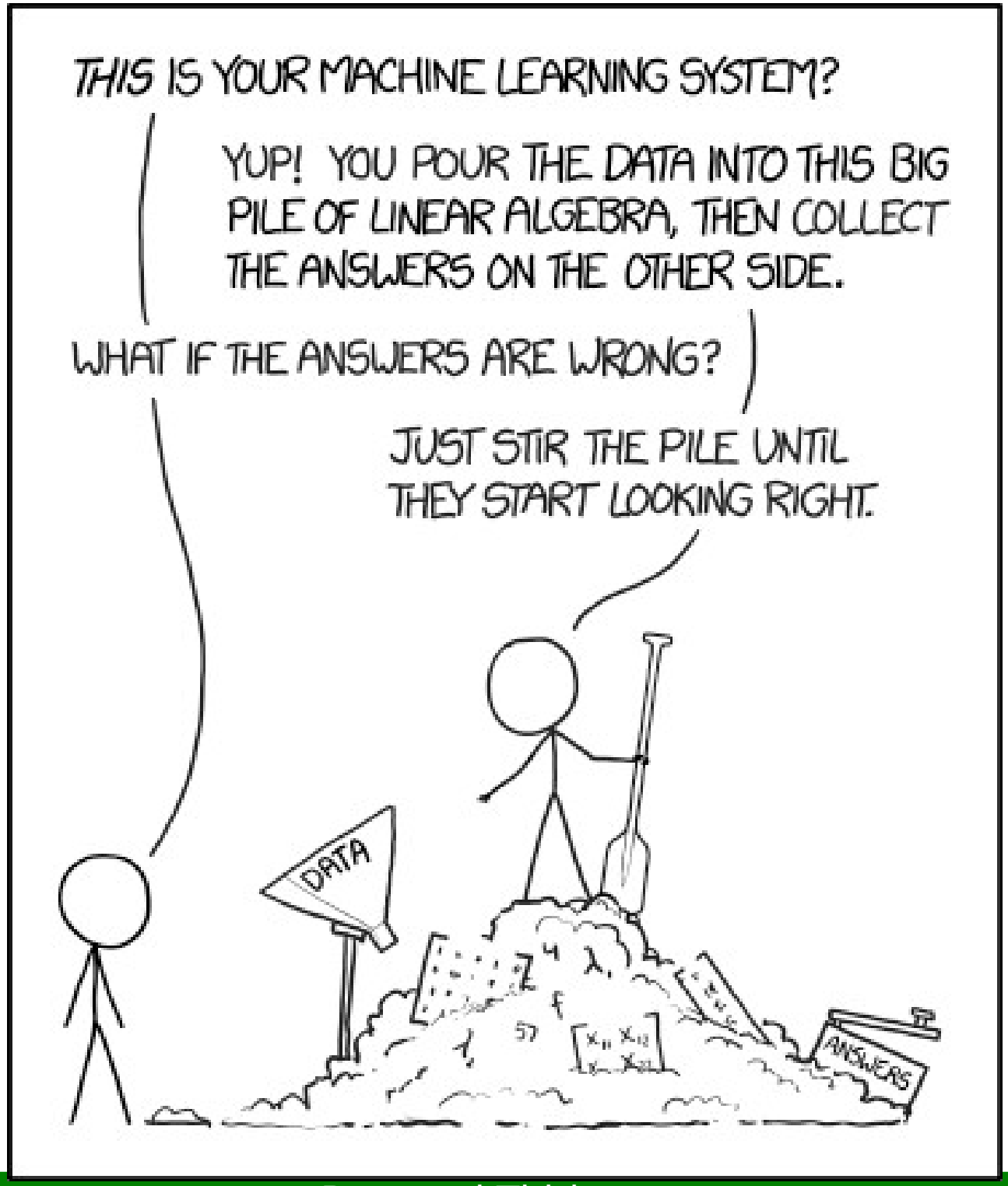
- Comprendre les concepts fondamentaux du machine learning… pour les démystifier
- Prendre avec critique les sorties des systèmes d'IA, comprendre leurs biais
- Comprendre comment les GAFAMs (et autres) aspirent les données

Learning Robots

startup créée par Thomas Deneux, Saclay
Robot AlphAI

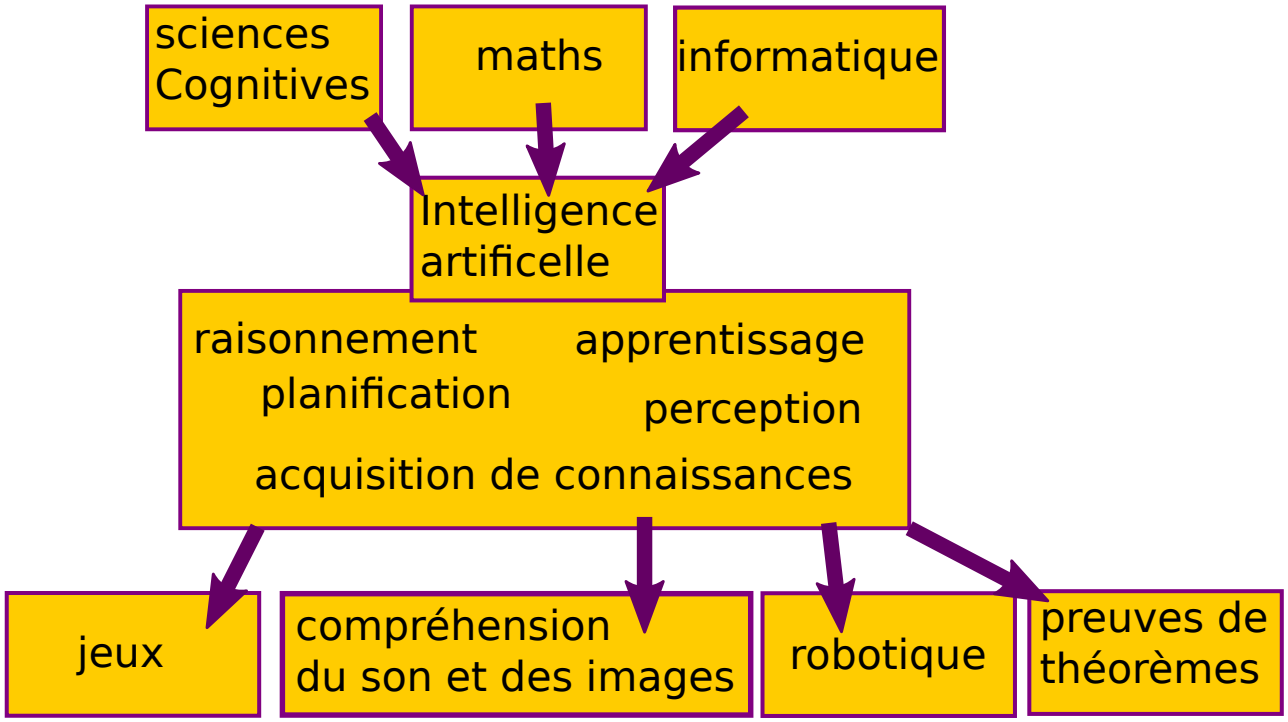
Conclusion

- L'IA a un très fort potentiel pour nous assister
  - Systèmes performants en moyenne
  - Capables de se baser sur bcp de données
- Limites difficiles à cerner !
- Besoin de réglementation des usages de l'IA
- Nous sommes **manipulables**: prendre conscience de nos modes de raisonnement
- Apprendre à gérer nos traces numériques
- Coût environnemental de l'IA


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

“Intelligence artificielle”

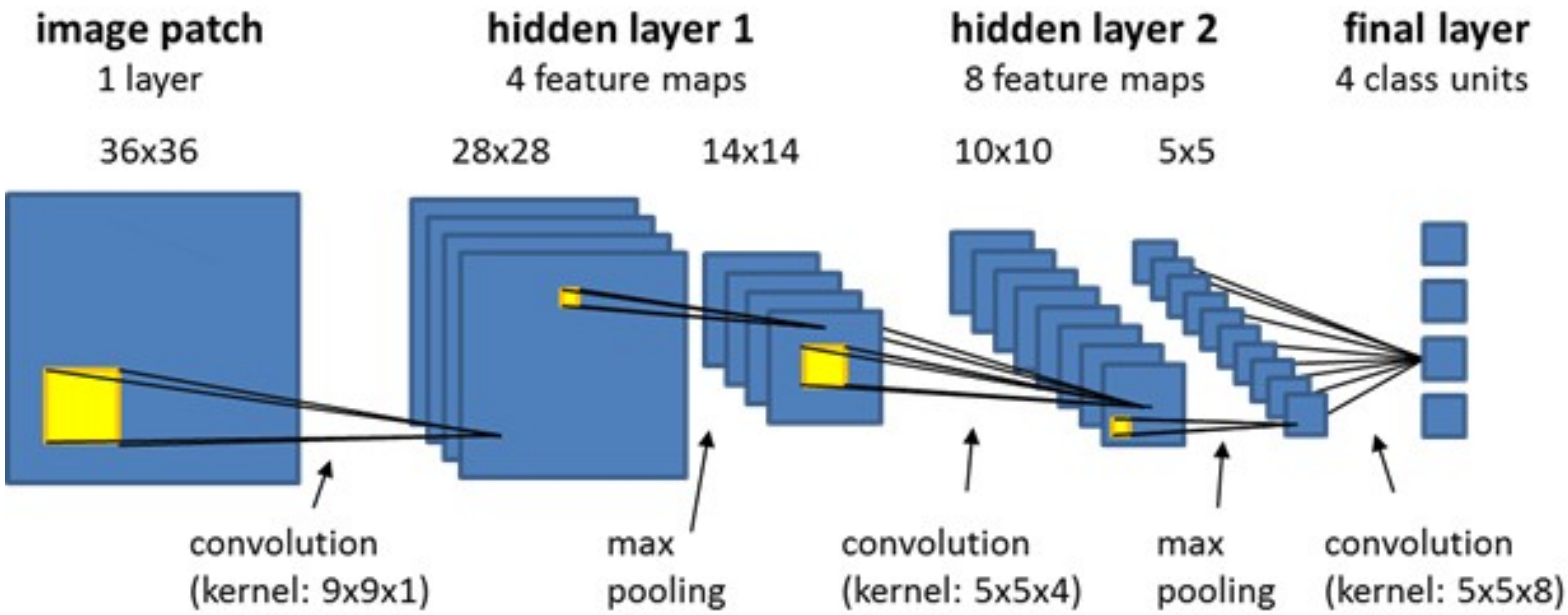
“l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de **simuler** l'intelligence humaine” Larousse



https://blog.ifs.com/2018/04/artificial-intelligence-defining-and-explaining-ai-related-terminology

La vision

- **Réseaux convolutionnels** pour la reconnaissance d'objet
  - invariance par translation
  - Robustesse aux petites déformations

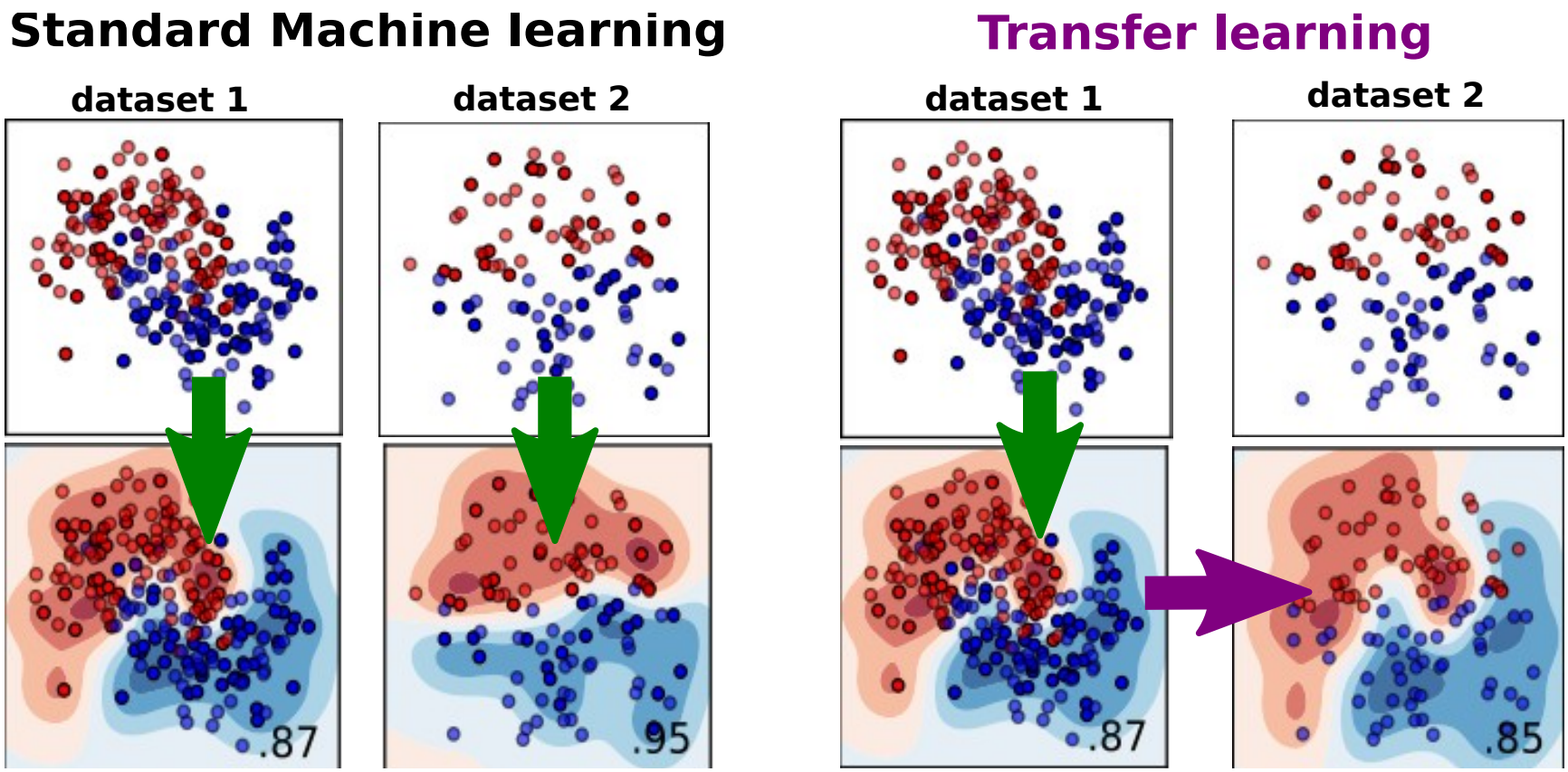


https://docs.ecognition.com/eCognition_documentation/User%20Guide%20Developer/8%20Classification%20-%20Deep%20Learning.htm

Le véhicule autonome

- Gérer un véhicule (trajectoire, sécurité)
- Peuvent réduire d'un facteur 10 les accidents
- Robustesse dans les situations critiques (insertion, imprévus) ?
- Stopé pour des questions de droit et d'assurance

A quoi l'apprentissage est-il bon ?



Apprentissage par transfert

- Généralisation forte
- Permet d'identifier des liens nouveaux