

Bases de l'apprentissage automatique

3DS – 2024

Direction de l'innovation et des relations avec les entreprises

cnrs formation
entreprises

L'intervenant



Alexandre Boucaud

- Ingénieur de recherche
au laboratoire
Astroparticule & Cosmologie
 - expertise en traitement
de gros volumes de données
astrophysiques
 - Pythonista averti
-

Programme

- **Lundi – 9h - 12h30**
 - introduction
 - apprentissage non supervisé
 - TP1 – visualisation de données
- **Mardi – 9h - 12h30**
 - apprentissage supervisé
 - TP2 – traitement des données numériques
- **Mercredi – 9h - 12h30**
 - pré-traitement des données
 - TP3 – traitement combiné données numériques et catégories
- **Jeudi – 9h - 12h30**
 - évaluation des modèles
 - optimisation d'hyper-paramètres
 - TP4 – hyperopt et cross validation
- **Vendredi – 9h - 17h**
 - réseaux de neurones / convolutifs
 - TP de deep learning

Modalités

- alternance de cours, cours TP et TP encadrés
- TP sur notebooks avec correction
- point en fin de séance pour passer en revue les questions / difficultés



Sommaire

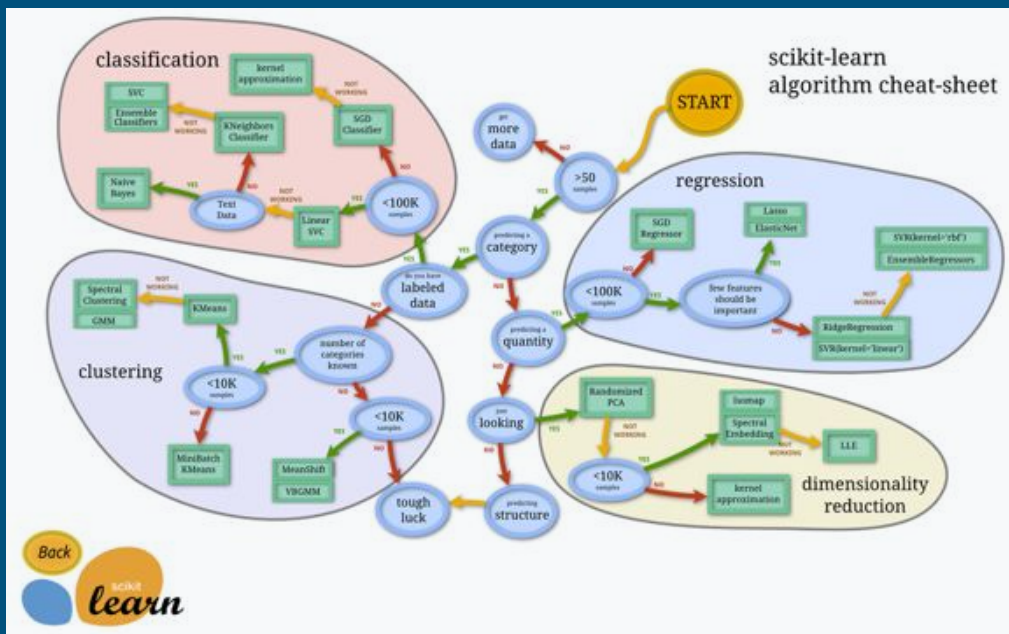
1. qu'est ce que le ML ?
2. d'où vient sa popularité ?
3. panorama des usages actuels
4. ce que vous allez apprendre dans ce cours



1. Qu'est ce que le ML ?

Le machine learning c'est..

un ensemble d'**algorithmes**



Liste non exhaustive :

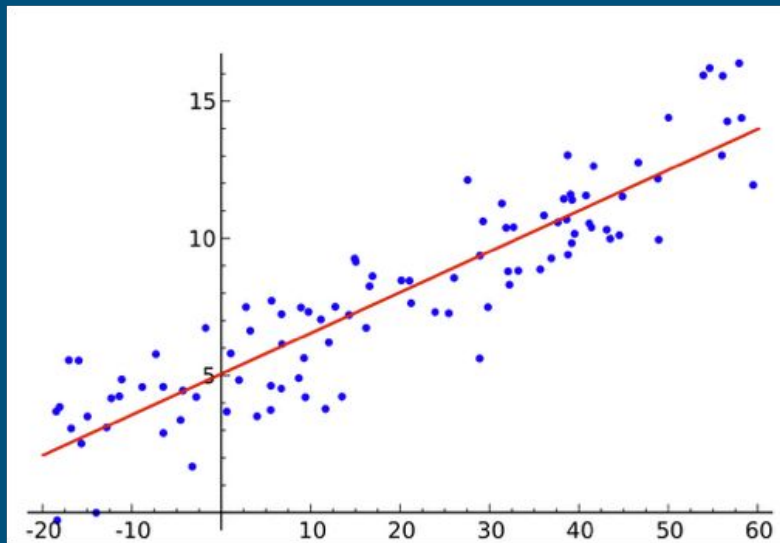
- Random Forests
- Nearest Neighbours
- Support Vector Machines
- (Deep) Neural Networks
- KMeans
- Gaussian Mixture Models
- Principal Component Analysis

Le machine learning c'est..

un ensemble d'**algorithmes**

qui produisent une **tâche précise**

régression

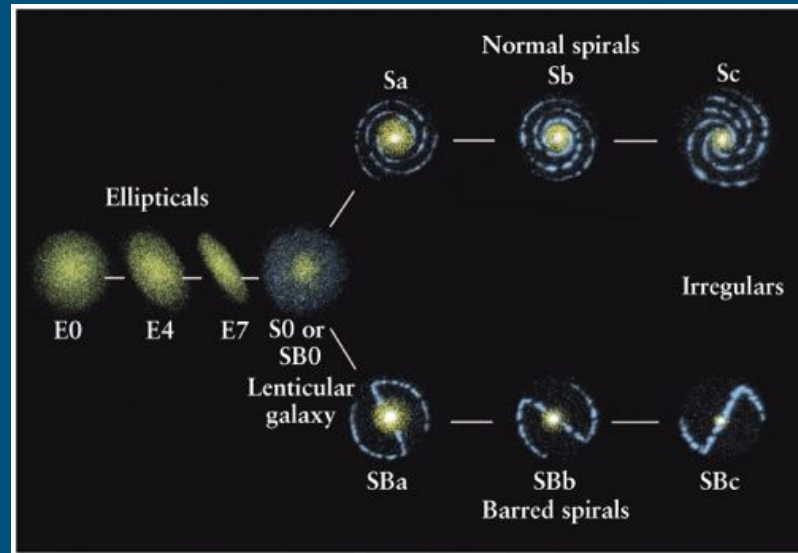


Le machine learning c'est..

un ensemble d'**algorithmes**

qui produisent une **tâche précise**

classification

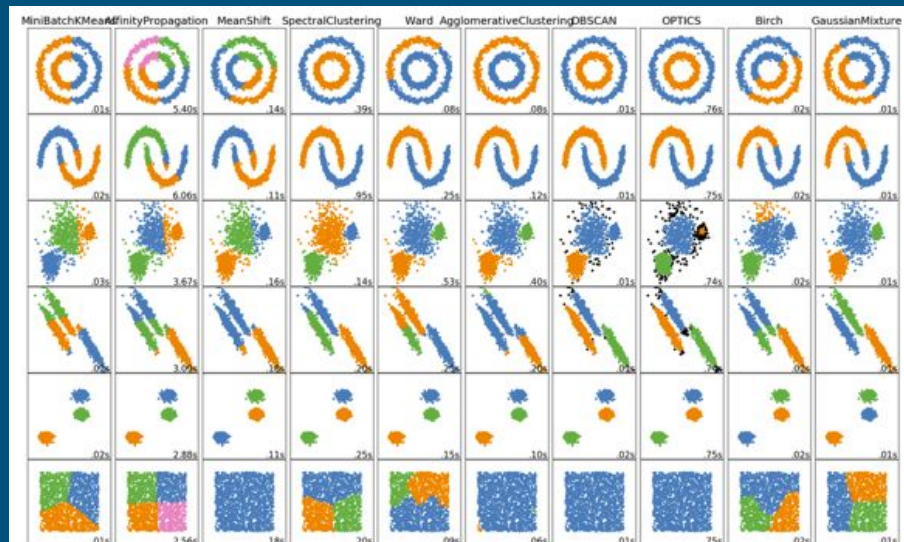


Le machine learning c'est..

un ensemble d'**algorithmes**

qui produisent une **tâche précise**

clustering



Le machine learning c'est..

un ensemble d'**algorithmes**

qui produisent une **tâche précise**

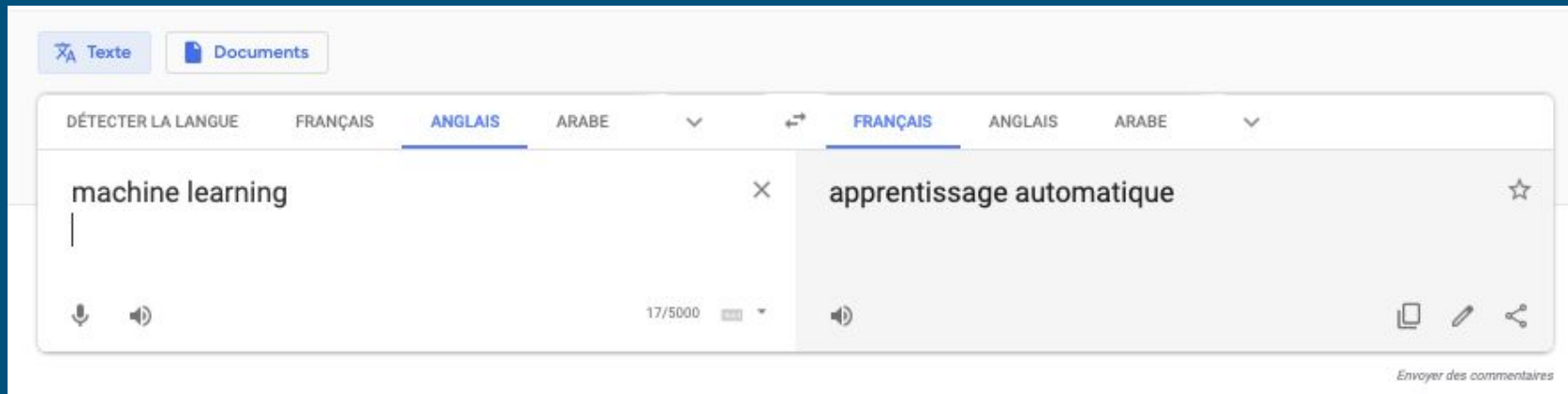
détection d'objets



Le machine learning c'est..

un ensemble d'**algorithmes**

qui produisent une **tâche précise**



Le machine learning c'est..

un ensemble d'**algorithmes**

qui produisent une **tâche précise**

avec des données

$$y = f(x)$$

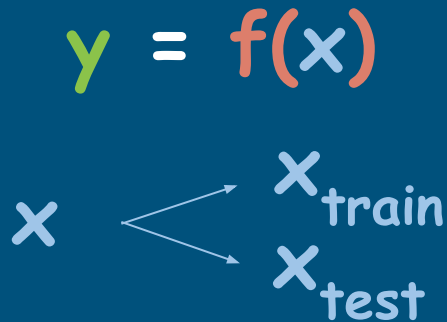
Autrement dit..

le machine learning fait
de l'**inférence** à partir de données

$$y = f(x)$$

Dans la pratique

1. séparation des données



2. entraînement

$$f(x_{\text{train}}) \rightleftharpoons y_{\text{train}}$$

3. évaluation

$$f(x_{\text{test}}) = ? y_{\text{test}}$$

Deux familles principales

Apprentissage supervisé

> action prédictive

permet de **répondre à une question** spécifique (régression, classification, etc)

nécessite des exemples où la solution est **connue** pour entraîner l'algorithme

se comporte comme de l'interpolation

Apprentissage non-supervisé

> action descriptive

permet de **découvrir des structures** dans les données (clustering, réduction de dimensions)

ne nécessite pas de données labellisées pour entraîner l'algorithme

peut s'utiliser **en amont de la classif/reg.**

Apprentissage supervisé (*supervised*)



problème de reconnaissance des
chiffres à partir d'images
(base de données MNIST)

<http://yann.lecun.com/exdb/mnist/>

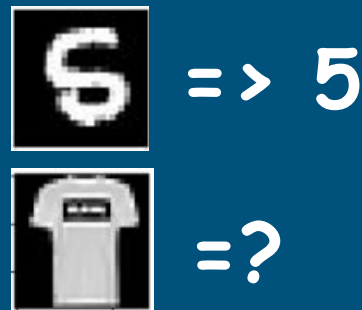
Apprentissage supervisé (*supervised*)



Entrainement :



Test :



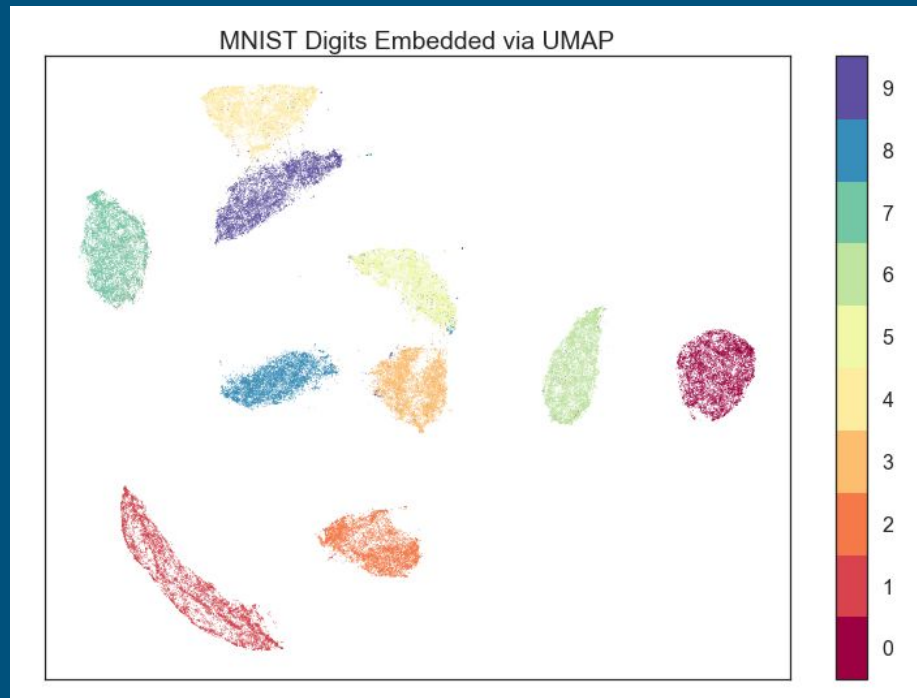


que dira un algorithme entraîné sur des images de chihuahua ?

Apprentissage non-supervisé (*unsupervised*)

visualisation des images MNIST
dans un espace 2d
avec l'algorithme UMAP

<https://github.com/lmcinnes/umap>



Ce qu'il faut retenir..

le machine learning c'est une **boîte à outils** pour les chercheurs

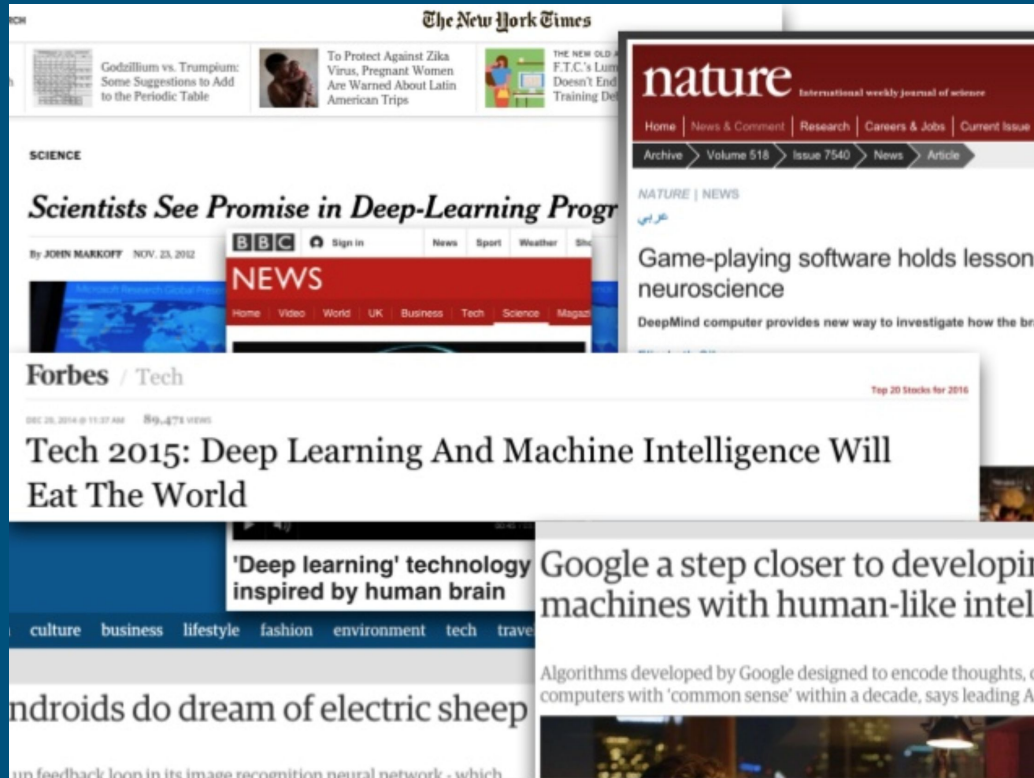
il existe une grande variété d'algorithmes et de modèles mais chacun d'entre eux est **plus adapté** pour une **tâche précise**

résoudre un problème avec du ML est surtout une question organisationnelle

- définition du problème
- caractérisation et pré-traitement des données

2. D'où vient sa popularité ?

Depuis quelques années, le machine learning est partout



mais il est surtout plébiscité par le terme IA

Quel est le lien entre intelligence artificielle et machine learning ?

Tout dépend de qui en parle..

les chercheurs détestent l'expression
"intelligence artificielle"

les industriels adorent



IA, la révolution n'a pas encore eu lieu - Michael Jordan (2018)

De quand datent ses premières utilisations ?

1957-69
dawn

techniques / tricks

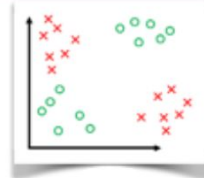
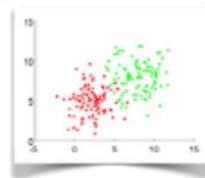
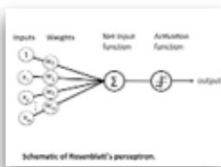
hardware

data

perceptron

early mainframes

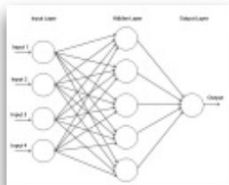
toy linear, small images, XOR



Golden age dans les années 90

1986-95
golden age

early NNs



workstations



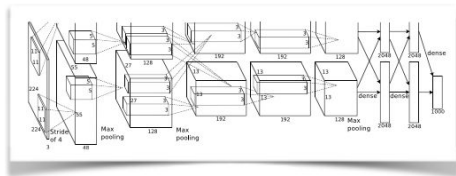
MNIST



Le renouveau après une période calme

2006-
deep learning

deep NNs



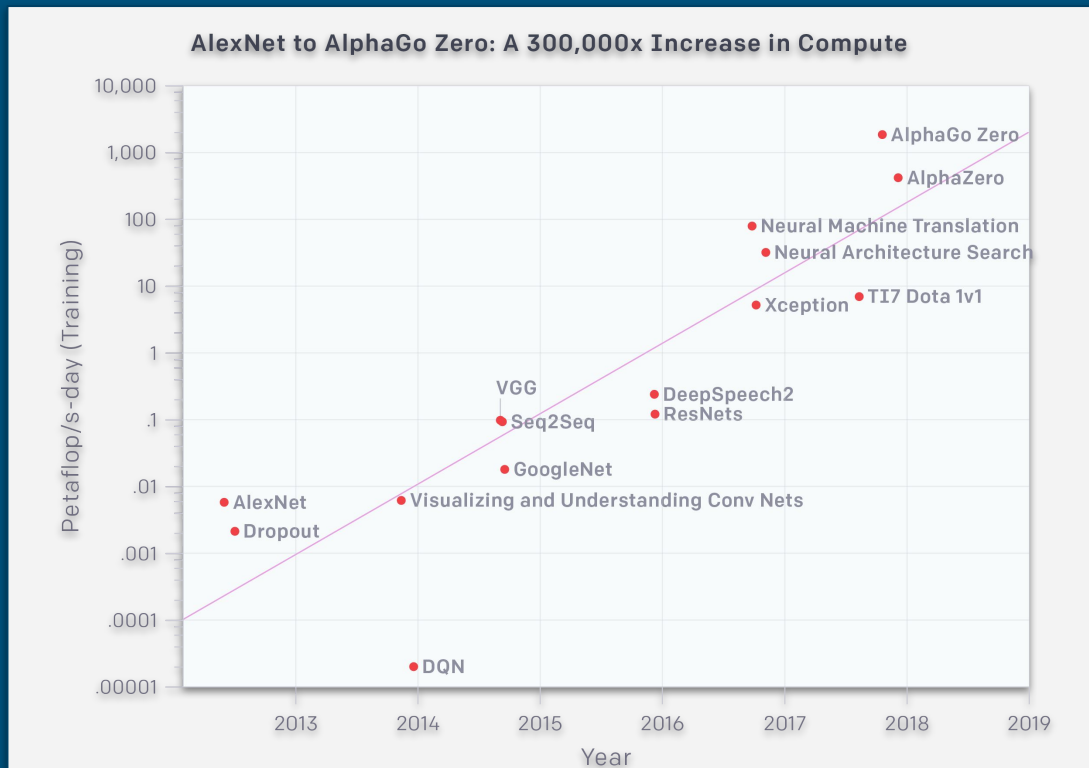
GPU, TPU, Intel Xeon Phi



Imagenet



Une utilisation sans cesse décuplée



les raisons du succès

- du **matériel** spécialisé de plus en plus performant
 - une grande disponibilité des **données**
 - une recherche stimulée sur les nouveaux **algorithmes**
 - un écosystème **open-source**
-

le matériel

hardware

- CPU dédiés
- GPU - graphics processing units
- TPU - tensor processing units



les données

data



les algorithmes

algorithms / models

Apprentissage..

- supervisé
- non-supervisé
- par récurrence
- par renforcement
- actif
- semi-supervisé
- ...



les logiciels libres

open-source software



3. Panorama des usages

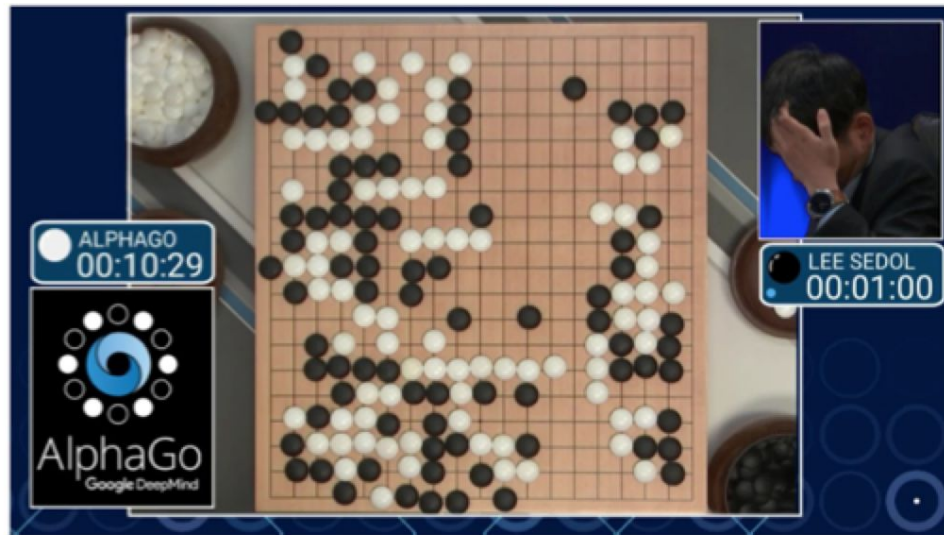
Parmi les applications courantes..

- traduction
 - du langage
 - image > texte
 - texte > image (génération)
- synthèse
 - de la parole
 - des images
 - des sons
- assistants personnels
- détections
 - reconnaissance des formes
 - reconnaissance faciale
 - aide à l'imagerie médicale
 - fraude (anomalies)
- voitures autonomes
- jeux

Examples



Style transfer - Gatys (2015)



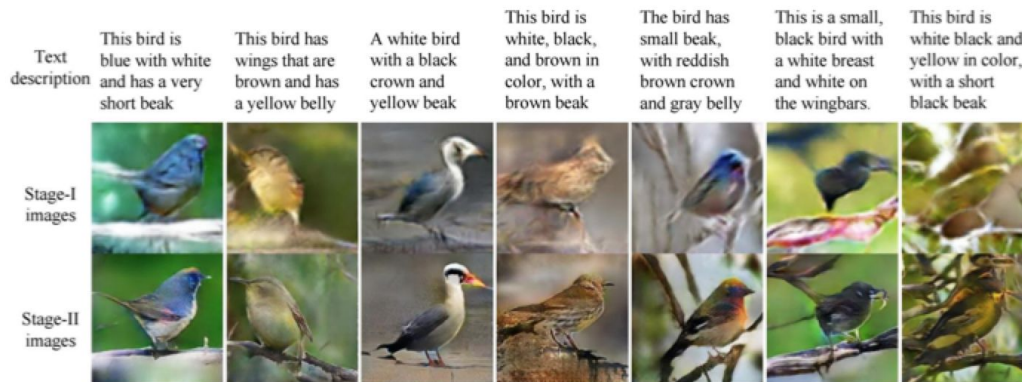
AlphaGo - DeepMind (2017)



Alexa - Amazon (2017)



Skin cancer diagnostic – Stanford (2017)

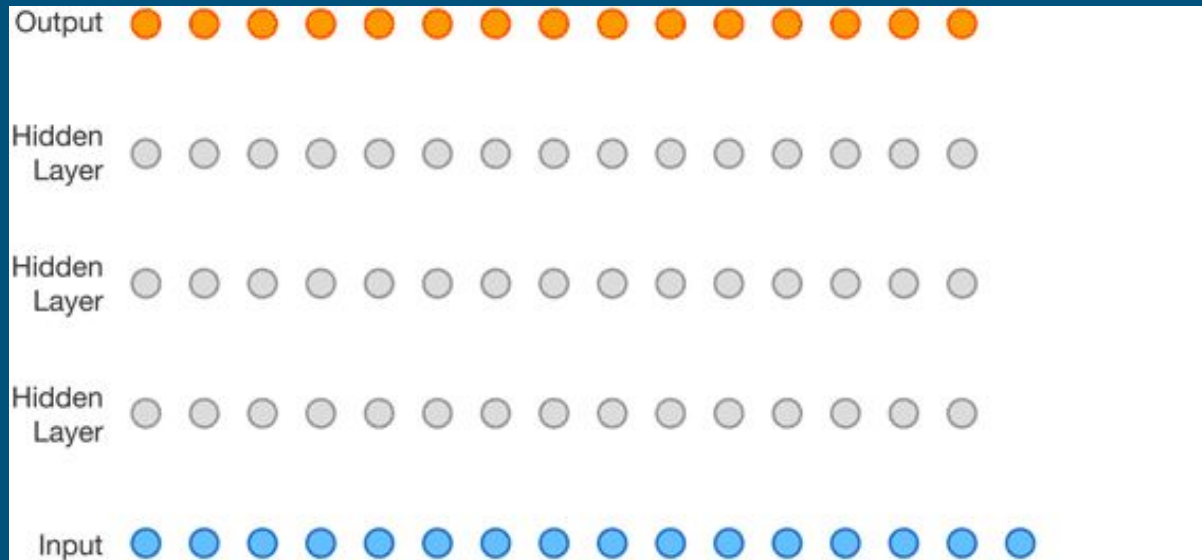


StackGAN v2 – Zhang (2017)

Colorisation d'image (2017)



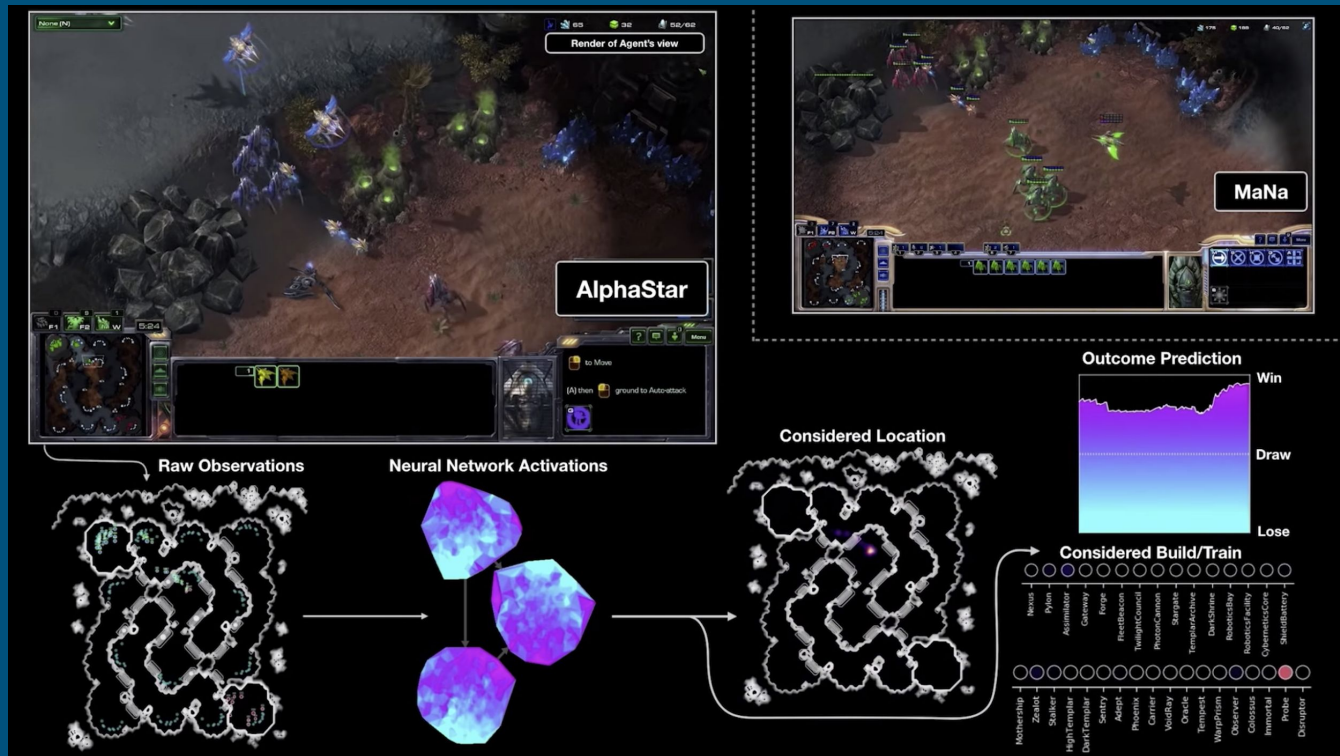
Synthèse de la parole (2017)



Synthèse des visages (2018)



Stratégie en temps réel (2019)



Efficient text parser/generator (2020)

GPT-3, l'intelligence artificielle qui a appris presque toute seule à presque tout faire

L'entreprise américaine OpenAI exploite le plus gros réseau de neurones artificiels au monde, effectuant une grande variété de tâches avec des résultats souvent bluffants, mais à la qualité imprévisible.

Le Monde (nov 2020)

<https://beta.openai.com>

[https://lacker.io/ai/2020/07/06/
giving-gpt-3-a-turing-test.html](https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test.html)

Image generation from text (2021)

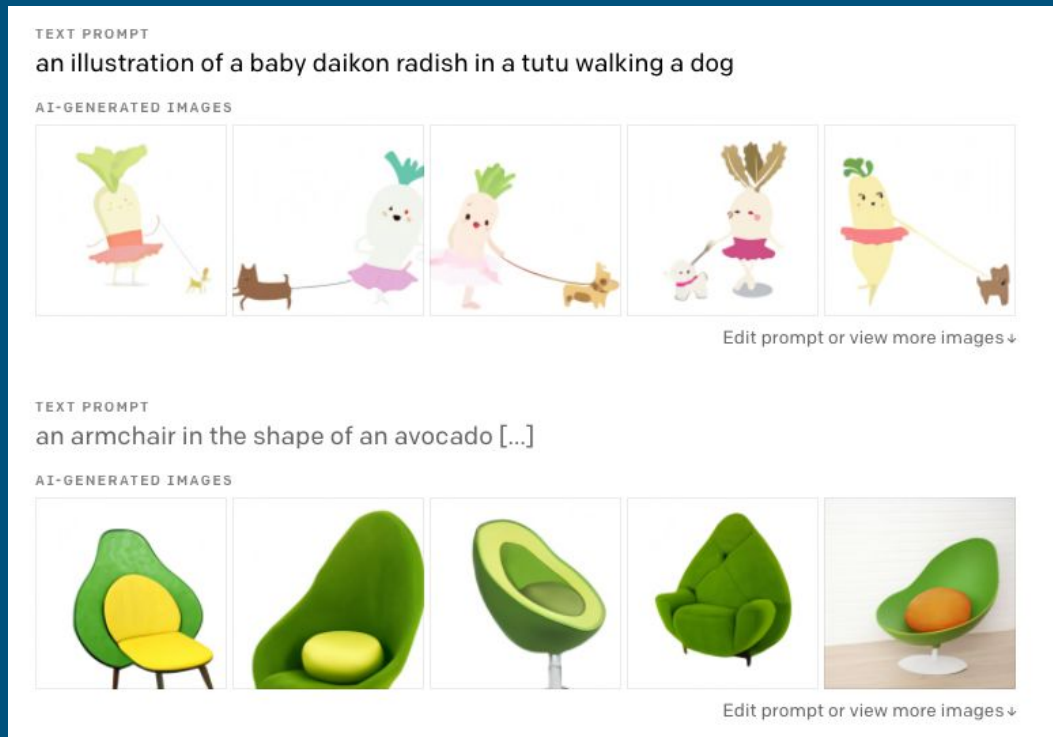


Image generation from text (2022)

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 vaporwave style as pixel art in a
photorealistic style



TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup
mixing sparkling chemicals as mad
scientists shopping for groceries working
on new AI research
as kids' crayon art on the moon in the
1980s underwater with 1990s technology



<https://openai.com/dall-e-2/>

ChatGPT (automne 2022)

<https://openai.com/blog/chatgpt/>



Can you describe what "modelisation" means in physics ?



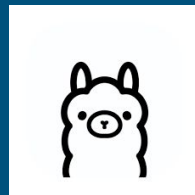
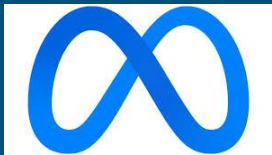
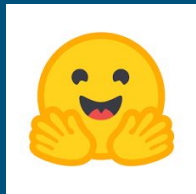
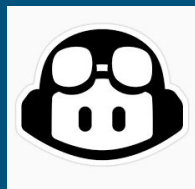
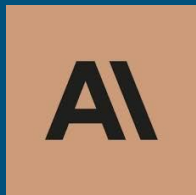
In physics, "modeling" refers to the process of creating a mathematical representation of a physical system or phenomenon. This representation, known as a model, is used to predict and explain the behavior of the system or phenomenon. The model can be a simple equation, a set of equations, or a computer simulation. The goal of modeling is to accurately represent the system or phenomenon being studied, and to make predictions about its behavior.

Image generation (2024)

<https://www.midjourney.com>



The LLM ecosystem in 2024



ChatGPT example

1



Can you describe what "modelisation" means in physics ?

2



How is this linked to machine learning ?

3



Could you summarise and translate to French please ?



La modélisation en physique consiste à créer une représentation mathématique d'un système physique ou d'un phénomène. Ce modèle est utilisé pour prédire et expliquer le comportement du système ou du phénomène. Les algorithmes d'apprentissage automatique peuvent être utilisés pour créer des modèles pour les systèmes physiques ou les phénomènes, et les modèles physiques peuvent être utilisés pour concevoir des algorithmes d'apprentissage automatique. De plus, l'apprentissage automatique peut être utilisé pour améliorer la précision des modèles physiques.

Résumé en français: La modélisation en physique consiste à créer des représentations mathématiques pour comprendre et prédire les comportements des systèmes physiques. Elle est liée à l'apprentissage automatique car ces derniers peuvent utiliser des modèles pour faire des prédictions à partir de données et améliorer la précision des modèles physiques.



Modeling in physics is closely linked to machine learning in several ways. One way is that machine learning algorithms can be used to create models for physical systems or phenomena. For example, a machine learning algorithm can be trained on a dataset of observations of a physical system, and then used to make predictions about the system's behavior.

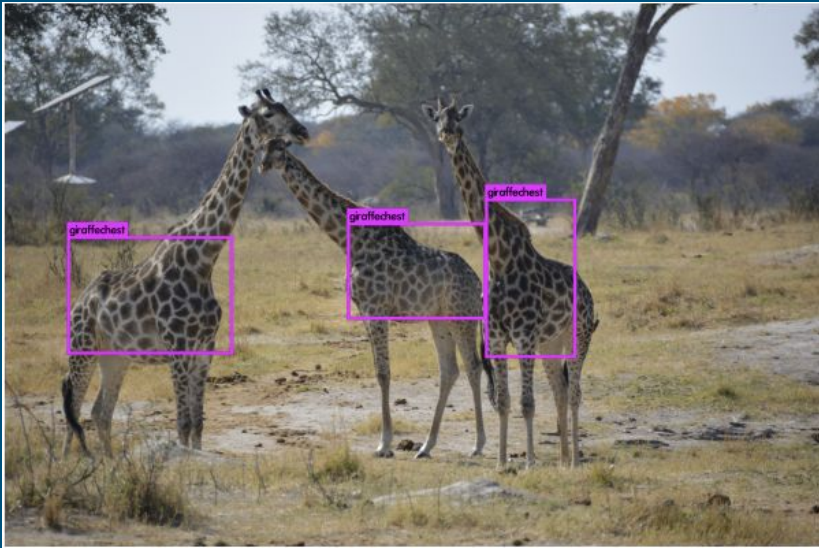
Another way is that physical models can be used as a basis for designing machine learning algorithms. For example, in physics, we might have a mathematical model that describes the dynamics of a system, and we can use this model as a guide to design a machine learning algorithm that can accurately simulate the system's behavior.

Additionally, Machine learning can be used to improve the accuracy of physical models by inferring the parameters of the model from experimental data, or by identifying patterns in the data that are not captured by the model.

In summary, the link between modeling in physics and machine learning is the use of mathematical representations to describe the behavior of physical systems. While physics uses mathematical models to understand the underlying laws of nature, machine learning uses mathematical models to make predictions from data. Both fields rely on the use of mathematical models, and both can inform and benefit from each other.

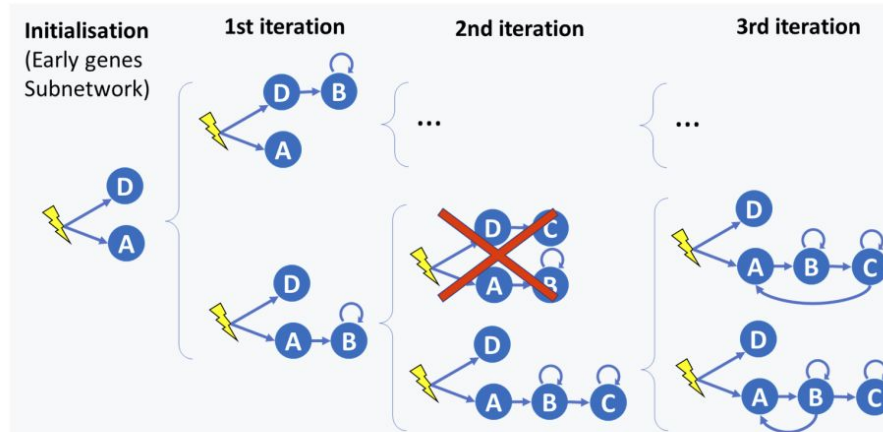
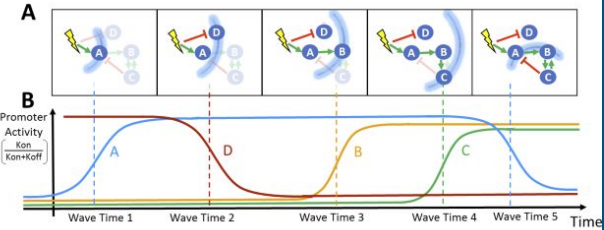
et en sciences..

Détection et suivi de girafes

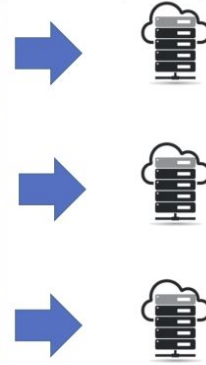


Prédiction de mutations génétiques

WASABI SPLITS & PARALLELIZES GRN INFERENCE PROBLEM

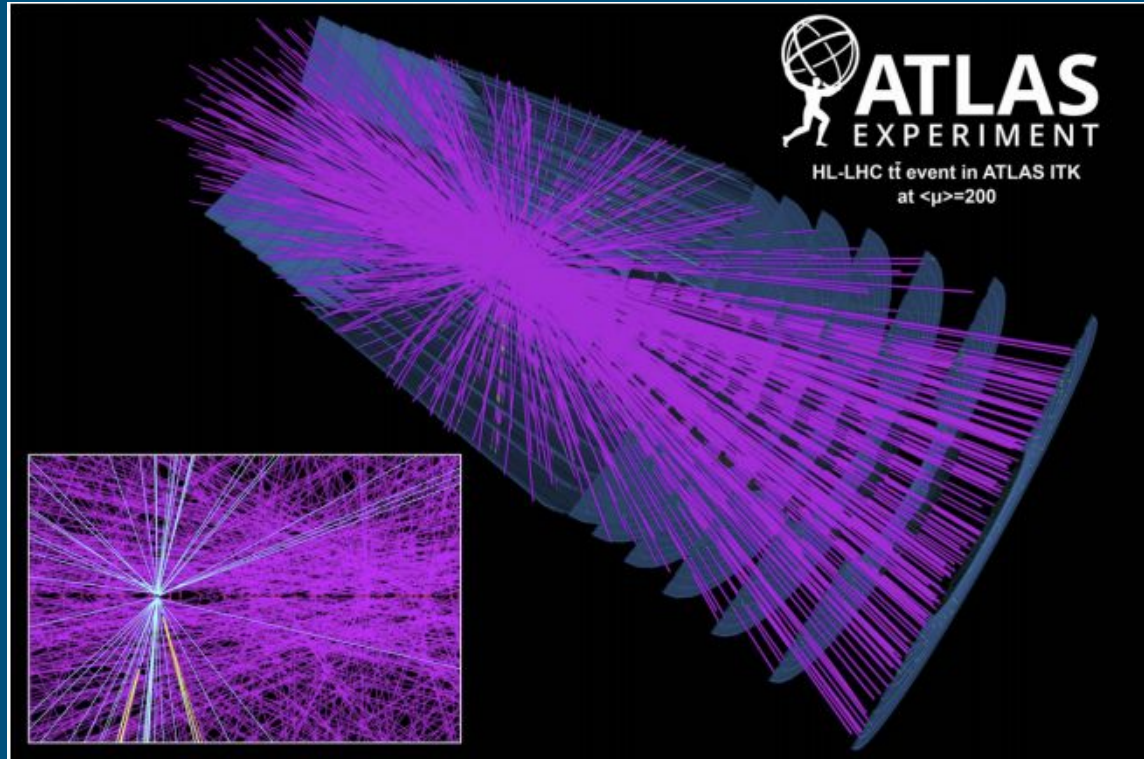


Cloud computing
(1000 cores / 15 days)

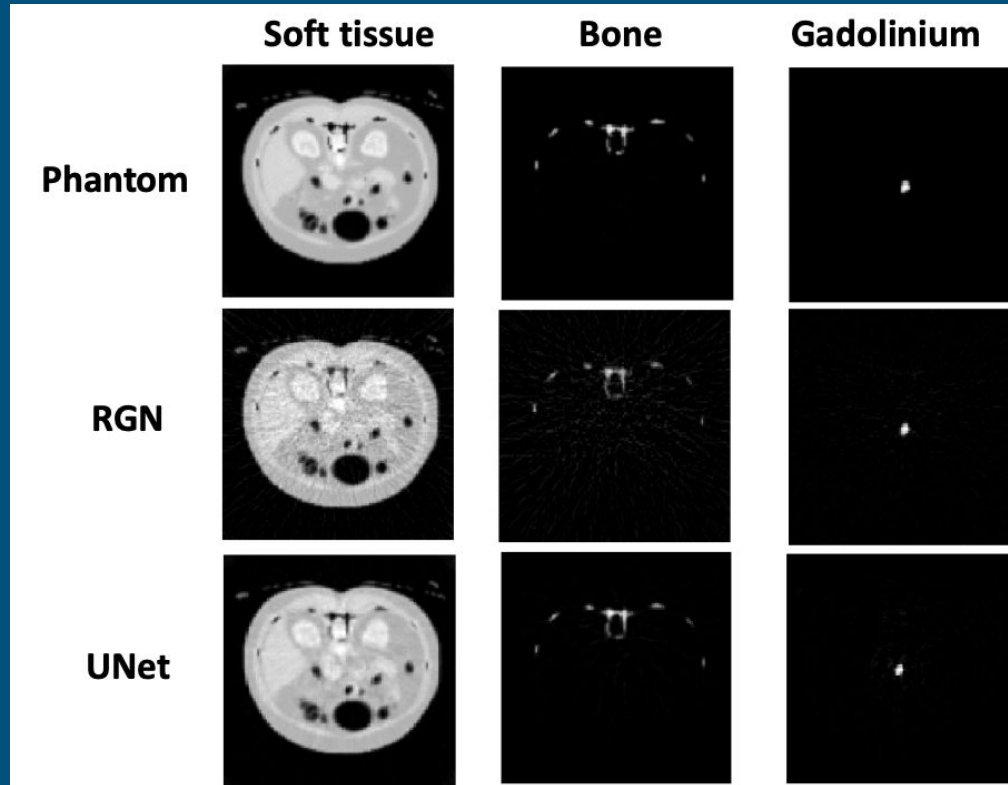


+ 
Machine Learning

Détection d'anomalies dans les calorimètres



Reconstruction et débruitage d'image méd.



Résoudre des équations et intégrales

DEEP LEARNING FOR SYMBOLIC MATHEMATICS

Guillaume Lample*
Facebook AI Research
glample@fb.com

François Charton*
Facebook AI Research
fcharton@fb.com

ABSTRACT

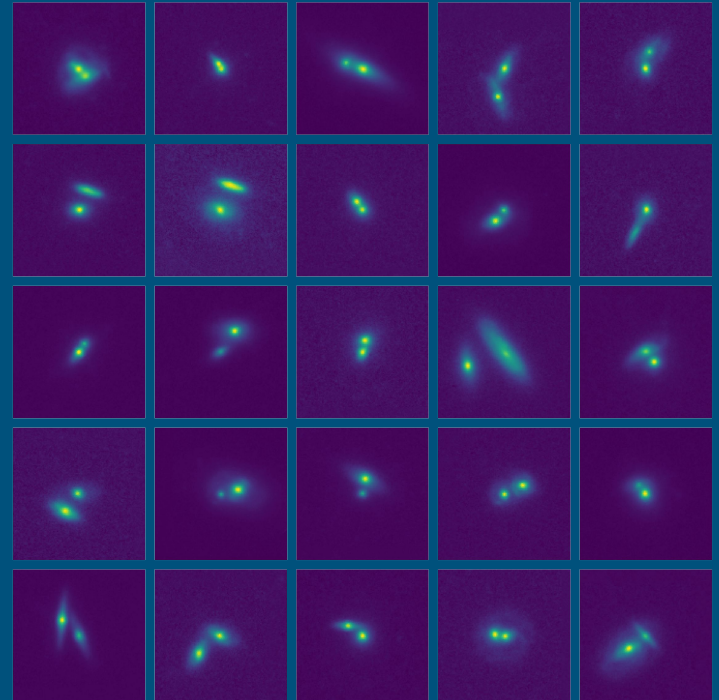
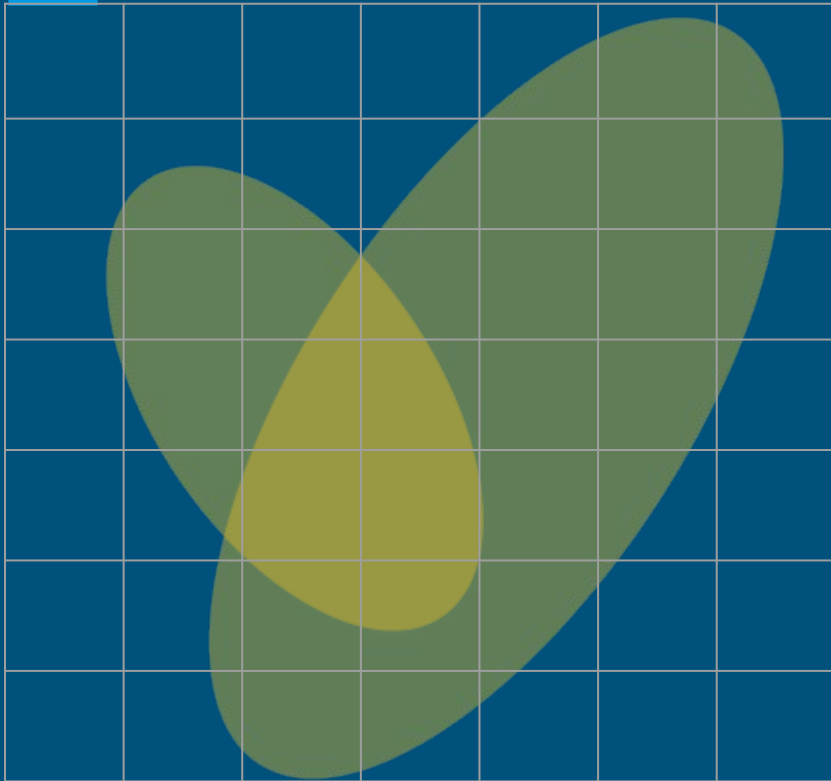
Neural networks have a reputation for being better at solving statistical or approximate problems than at performing calculations or working with symbolic data. In this paper, we show that they can be surprisingly good at more elaborated tasks in mathematics, such as symbolic integration and solving differential equations. We propose a syntax for representing mathematical problems, and methods for generating large datasets that can be used to train sequence-to-sequence models. We achieve results that outperform commercial Computer Algebra Systems such as Matlab or Mathematica.

Simulation de galaxies

A vous de trouver l'image simulée..

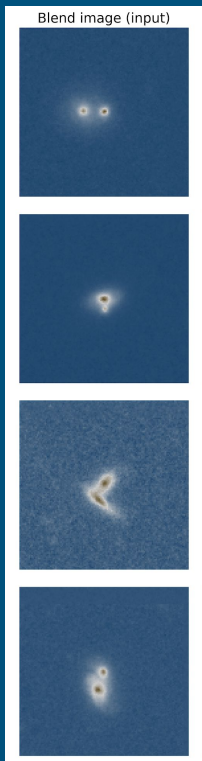


Séparation de galaxies (mon sujet principal)

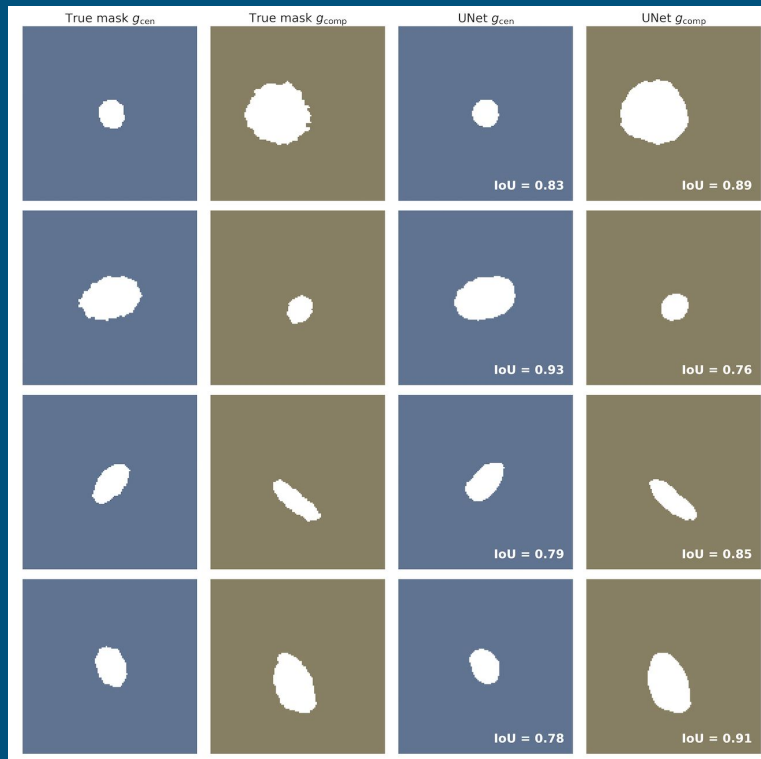


Séparation de galaxies

INPUT IMAGES
(TEST SET)



TRUE
SEGMENTATION

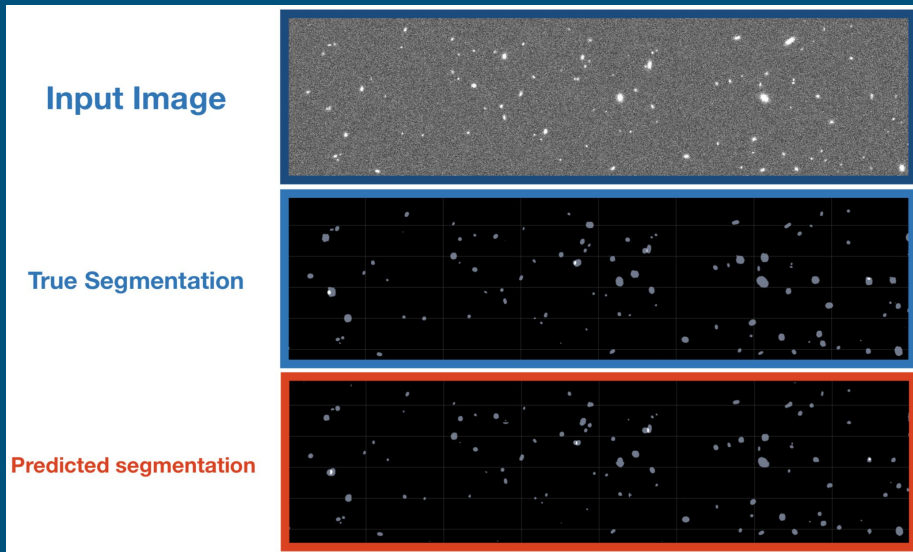
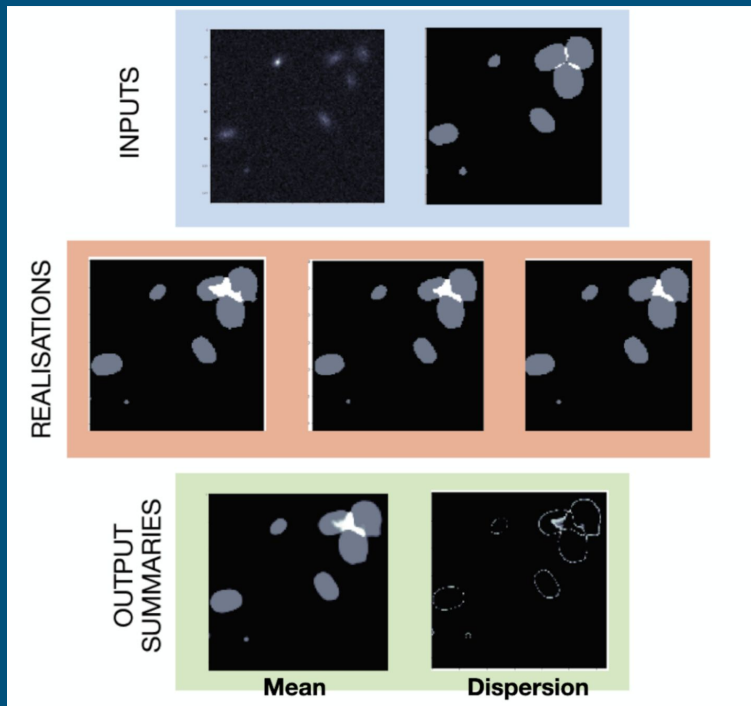


PREDICTED
SEGMENTATION

Séparation probabiliste



Hubert Bretonnière
thèse soutenue en sept 2022



4. Ce que vous allez apprendre

Vue d'ensemble du programme

Introduction au machine learning

Non supervisé

clustering, visualisation et réduction de dimensions

Supervisé

régression et classification

Chaîne de processing

prétraitement, sélection des modèles

Réseaux de neurones et deep learning

Réseaux de neurones

théorie

perceptron multi-couche

Deep learning

traitement d'images avec réseaux de convolution