Avancées et réalités de l'IA

CentraleSupélec | université PARIS-SACLAY

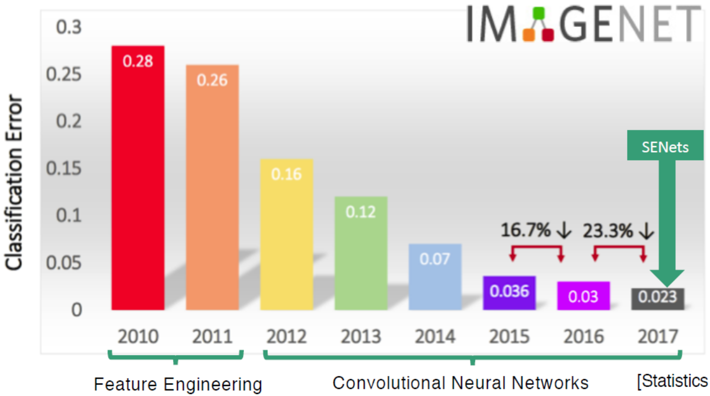
September 8, 2023

IA: réalités

Partout !


Faces
Flowers
Hand
Monitoring
Robotics
Security
Medical
Sport
CAB/SEDAN
BIKE
Transport

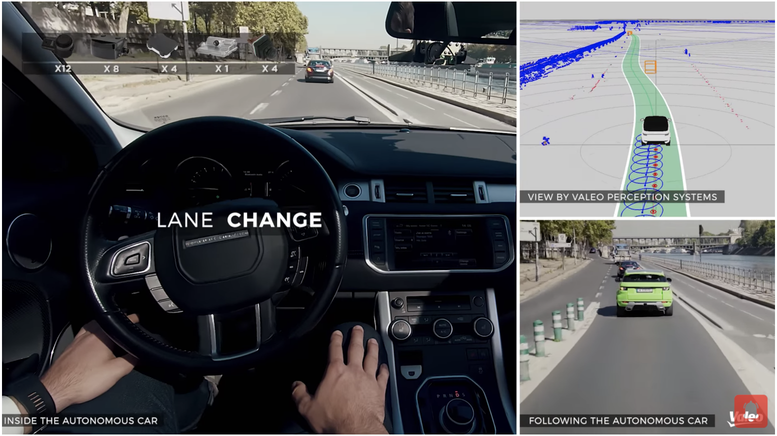
Très performante !



[Statistics provided by ILSVRC]

Plus performante que les humains ?

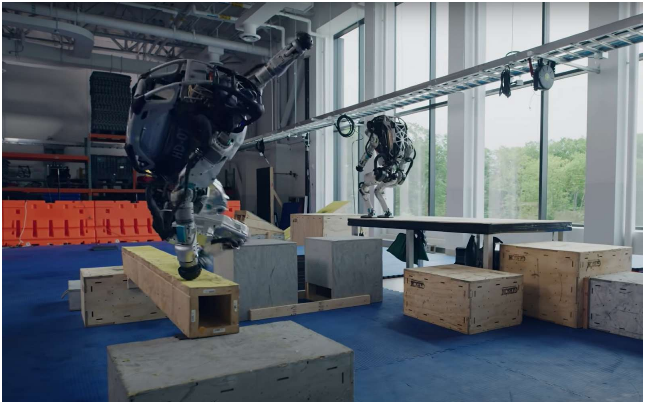
Capable de conduire dans les rues de Paris



Voir :`https://www.youtube.com/watch?v=9mBL16JuvsM`

Des robots qui pratiquent le parkour

Atlas, le robot de Boston Dynamics



https://www.youtube.com/watch?v=tF4DML7FIWk

Capable de générer des données très réalistes!

Figure 1: Test it here : `https://thispersondoesnotexist.com/`

Capable de générer du texte

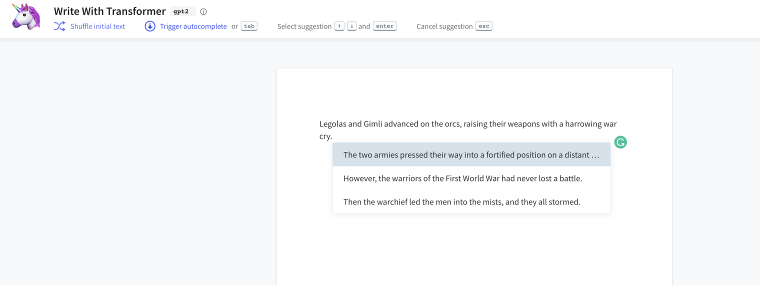
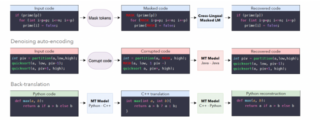
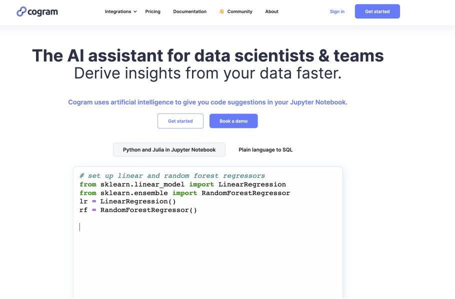


Figure 2: Test it here : `https://transformer.huggingface.co/`

Capable de générer du code, des programmes informatiques





https://arxiv.org/pdf/2006.03511.pdf et
https://www.cogram.com/

Capable de créer de l'art !

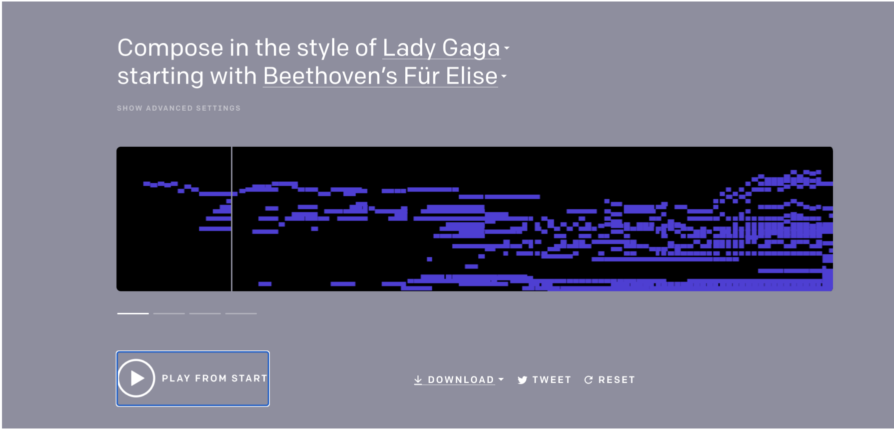


Figure 3: Test it here : `https://openai.com/blog/musenet/`

Capable de créer de l'art !

Figure 4: The next Rembrandt: `https://www.nextrembrandt.com/`

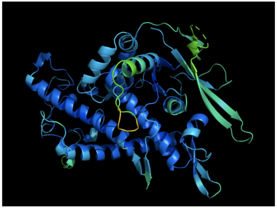
Ce qui pose des problèmes de propriété intellectuelle : qui est l'auteur ? (`https://cacm.acm.org/magazines/2020/7/245693-ai-authorship/fulltext`).

Capable de prédire la structure des protéines à partir de leur séquence en acides aminés

‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway



A protein’s function is determined by its 3D shape. Credit: DeepMind

An artificial intelligence (AI) network developed by Google AI offshoot DeepMind has made a gargantuan leap in solving one of biology’s grandest challenges — determining a protein’s 3D shape from its amino-acid sequence.

Capable de prédire la structure des protéines à partir de leur séquence en acides aminés

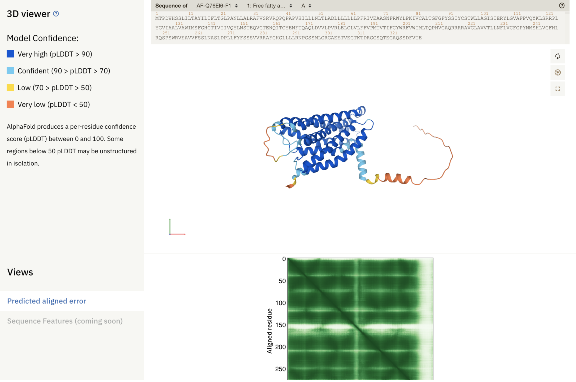


Figure 5: AlphaFold

https://alphafold.ebi.ac.uk/

Capable de prédire la structure des protéines à partir de leur séquence en acides aminés

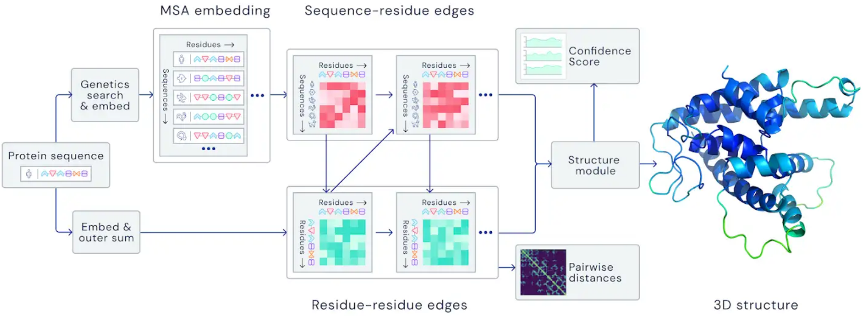


Figure 6: AlphaFold

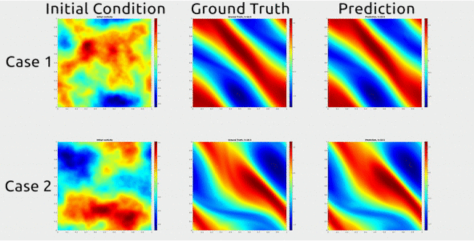
Capable de résoudre des équations aux dérivées partielles !

ARTIFICIAL INTELLIGENCE

AI has cracked a key mathematical puzzle for understanding our world

Partial differential equations can describe everything from planetary motion to plate tectonics, but they're notoriously hard to solve.

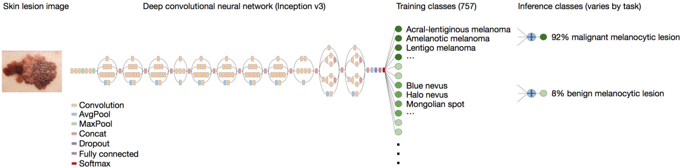
By Karen Hao October 30, 2020



IA: réalités

Plus performante que les humains ?

Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

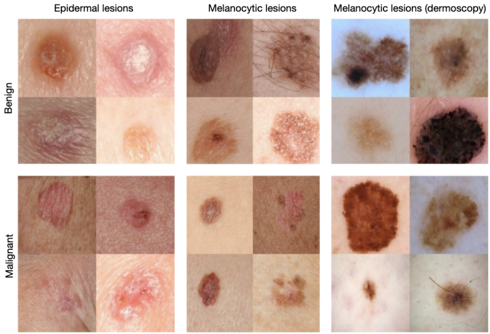


Dermatologist-level classification of skin cancer with deep neural networks - Esteva et al [Nature,2017]

Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

The task

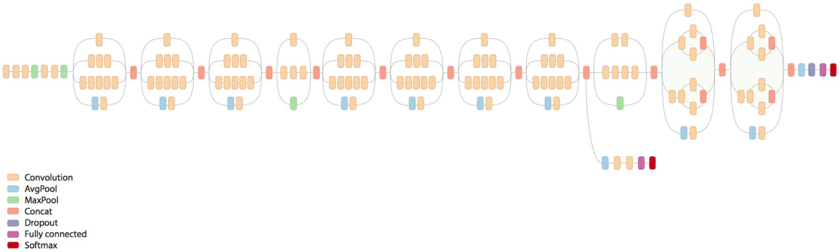
Skin cancer detection : a fine-grained visual recognition task (2,032 different diseases, fine-grained variability) but evaluation is done on two binary classification tasks (*keratinocyte carcinomas versus benign seborrheic keratoses; and malignant melanomas versus benign nevi*).



Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

The model

Inception v3 CNN architecture, pre-trained on ImageNet and fine-tuned on the target dataset[1]

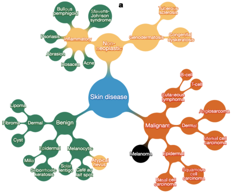


[1] https://ai.googleblog.com/2016/03/train-your-own-image-classifier-with.html

Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

The data

- 129,450 clinical images, including 3,374 dermoscopy images, **annotated by dermatologists**.
- Images are organized in a tree-structured taxonomy of 2,032 diseases, derived by dermatologists using a bottom-up procedure: individual diseases, initialized as leaf nodes, were merged based on clinical and visual similarity, until the entire structure was connected.
- Training dataset ; 127,463 training and validation images and testing dataset : 1,942 biopsy-labelled test images with no-overlap (same lesion, multiple viewpoints)

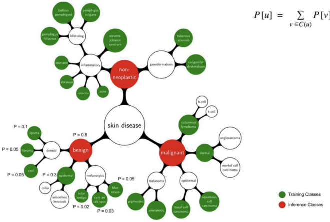


Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

Drom Disease to training classes
Disease partitioning algorithm : partitions individual diseases into training classes whose individual diseases are clinically and visually similar and with contraints on the size of the class (maxClassSize = 1,000) : **disease partition of 757 classes**.

Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

From training classes to inference classes.



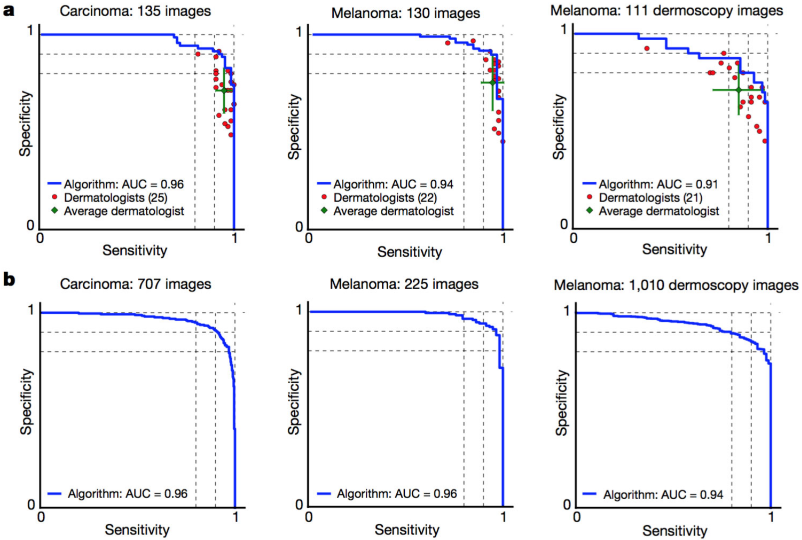
$$P[u] = \sum_{v \in C(u)} P[v]$$

Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]

Experimental protocol

- We test against 21 board-certified dermatologists on biopsy-proven clinical images.
- Two critical binary classification use cases:
    - malignant carcinomas versus benign seborrheic keratoses: identification of the most common cancers
    - malignant melanomas versus benign nevi : identification of the deadliest skin cancer

Dermatologist-level classification of skin cancer with deep neural networks [Esteva et al, 17]



Training accurate deep models

Usually requires lots of well-labeled data

- Data collection and annotation is expensive, tedious, time consuming
- Crowd-sourcing may be infeasible for proprietary data.
- For some tasks, data may not be available at all (long tail distribution)

Assuming the availability of large well-labeled dataset is not a realist assumption

Availability of labeled data

Synset: mushroom
Definition: any of various fleshy fungi of the subdivision Basidiomycota consisting of a cap at the end of a stem arising from an underground mycelium.
Popularity percentile:: 84%
Depth in WordNet: 7

Synset: mushroom
Definition: mushrooms and related fleshy fungi (including toadstools, puffballs, morels, coral fungi, etc.).
Popularity percentile:: 82%
Depth in WordNet: 8

Synset: mushroom
Definition: fleshy body of any of numerous edible fungi.
Popularity percentile:: 82%
Depth in WordNet: 6

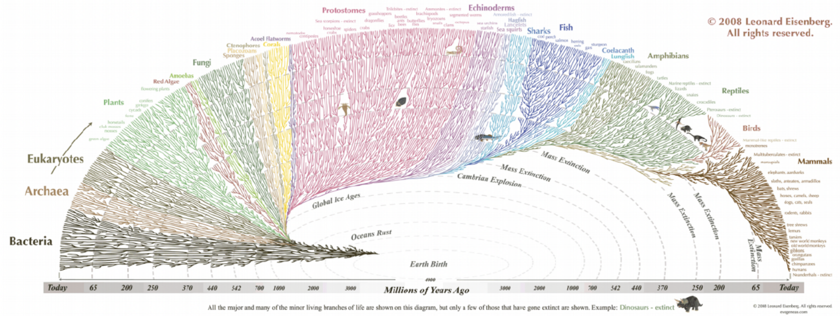
Synset: stuffed mushroom
Definition: mushrooms stuffed with any of numerous mixtures of e.g. meats or nuts or seafood or spinach.
Popularity percentile:: 69%
Depth in WordNet: 8

Synset: mushroom sauce
Definition: brown sauce and sauteed mushrooms.
Popularity percentile:: 69%
Depth in WordNet: 9

ImageNet has 30 mushroom synsets, each with ≈1000 images.

Slide credit: Christoph Lampert

Availability of labeled data



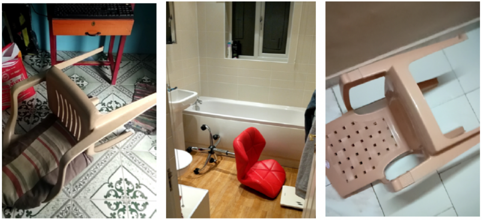
In nature, there are ≈14,000 mushroom species !

Availability of labeled data

Classes with many examples may still suffer from bias !

ImageNet chairs

- Few rotations/viewpoints
- Typical backgrounds

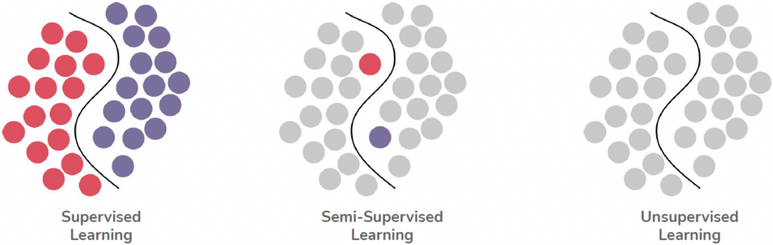


Images from the ObjectNet dataset [Barbu et al., NeurIPS 2019]

ImageNet classifier fails to detect these chairs

Learning with limited labeled data

Alternatives to supervised learning



Richness of unlabeled data

- Data privacy friendly
- Cheaper
- Scales better

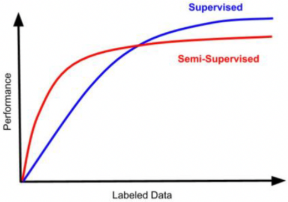
Author of the slides of this part : Y. Ouali, PhD student, MICS

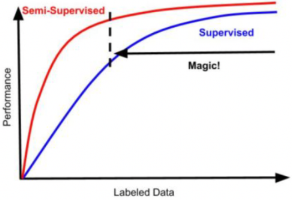
Learning with limited labeled data

Semi-supervised Learning

Semi-supervised Learning

> Semi-supervised learning (SSL) is halfway between supervised and unsupervised learning. In addition to unlabeled data, the algorithm is provided with some supervision information – but not necessarily for all examples. Often, this information will be the targets associated with some of the examples. In this case, the data set $X = (x_i)\,;i \in [n]$ can be divided into two parts: the points $X_l := (x_1, \ldots, x_l)$, for which labels $Y_l := (y_1, \ldots, y_l)$ are provided, and the points $X_u := (x_{l+1}, \ldots, x_{l+u})$, the labels of which are not known.
>
> — ***Chapelle et al.*** — *SSL book*





Reference : SSL book

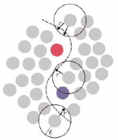
Semi-supervised Learning

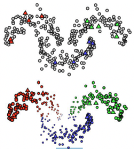
Main assumptions about the structure of data

- **The Smoothness Assumption**: if two input points that reside in a high-density region are close then so should be their corresponding outputs.
- **The Cluster Assumption** : If points are in a same cluster, they are likely to be of the same class (particular case of the Smoothness Assumption in which we suppose that input data points from cluters)
- **The Manifold Assumption** : The (high dimensional) data (roughly) lie on a low-dimensional manifold, i.e. discover the low dimensional representation from unlabeled data and use the labeled data to solve the simplified task in the lower dimensional space.

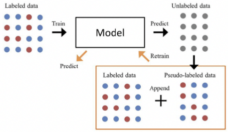
Main methods

- **Consistency training** : a realistic perturbation applied to the unlabeled point should not change significantly the prediction, i.e. the model can be trained to have a consistent prediction on a given unlabeled example and its perturbated version.
- **Pseudo-labeling methods** : leveraging of a model trained on a labeled set to produce additional training examples by labeling instances of the unlabeled set based on some heuristics.
- **Generative models** : If a generative model is able to generate realistic images from the data distribution $p(x)$, then it must learn transferable features to a supervised task $p(y|x)$
- **Graph-bases methods** : Labeled and unlabeled data are considered as nodes of a graph and the objective is to propagate the labeled nodes to the unlabeled ones using the similarity between nodes and edge information.





Graph based Label



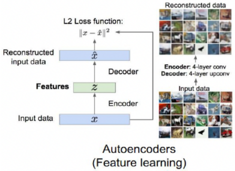
Pseudo labeling

Learning with limited labeled data

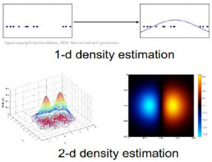
Unsupervised learning

In unsupervised learning, we are interested in extracting information from the unlabeled data x_i (without using labels y_i)

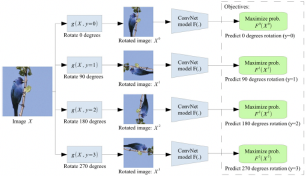
- Extracting low-dimensional embeddings of the data : **Autoencoders**.
- Generative models able to approximate the data distribution $P_X(x)$: **GANs : Generative adversarial networks**.
- Defining new training objectives based on the data itself to train the model : **self-supervision**.



Auto-encoder and dimensionality reduction

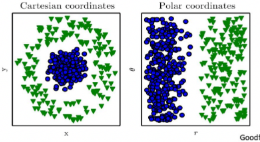


GaNs and density estimation

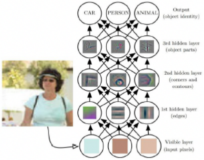


Self-supervision

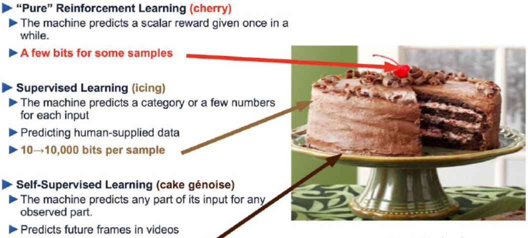
Motivation : **representation learning matters**. We can learn useful representations without any labels (the goal of self-supervised learning)



Goodfellow



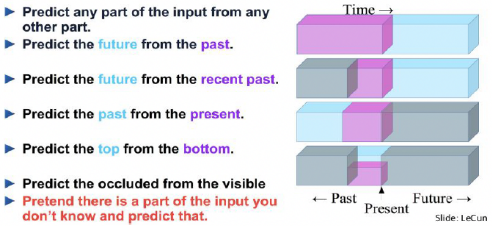
Goodfellow



Yann LeCun's cake

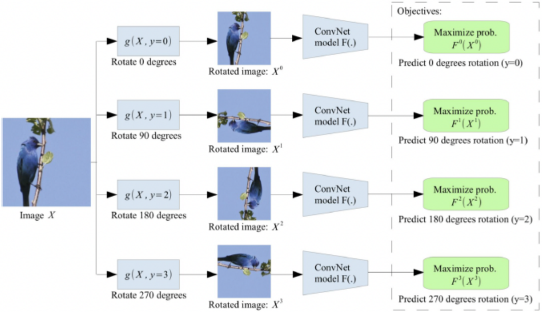
Unsupervised learning : what is self-supervision ?

- A form of unsupervised learning in which the data provides the supervision.
- In general, withhold some parts of the data and the goal is to predict it from the remaining parts : **Pre-text task**
- Goals :
  - Learn equally good features without supervision.
  - Deploy similar quality models without relying on too many labels.
  - Generalize potentially better because learn more about the world.



Self-supervised learning : examples

With image distorsions (rotations) : create a class by augmenting the same image multiple times with rotations. The model is trained to predict which rotation has been applied.



# Rotations	Rotations	CIFAR-10 Classification Accuracy
4	0°, 90°, 180°, 270°	**89.06**
8	0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°	88.51
2	0°, 180°	87.46
2	90°, 270°	85.52

Learning with limited labeled data

Few shot learning

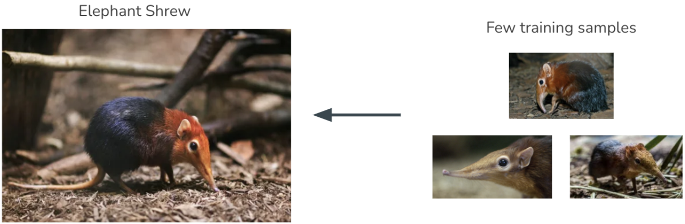
Few-shot learning

Equip the learner with ability to rapidly learn new concept with few training samples.

?????

Few-shot learning

The learner is only provided with few samples, for examples one or five images, in order to recognize this novel category (Elephant Shrew)

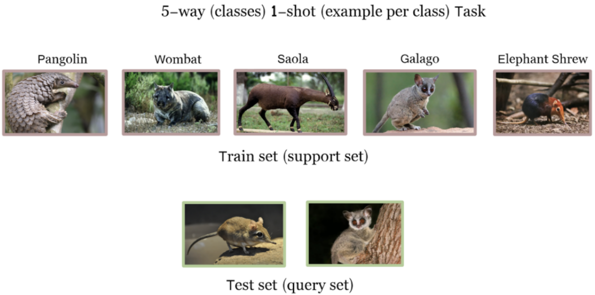


Few-shot learning

A few-shot classification task: $N-$way $K-$shot classification task

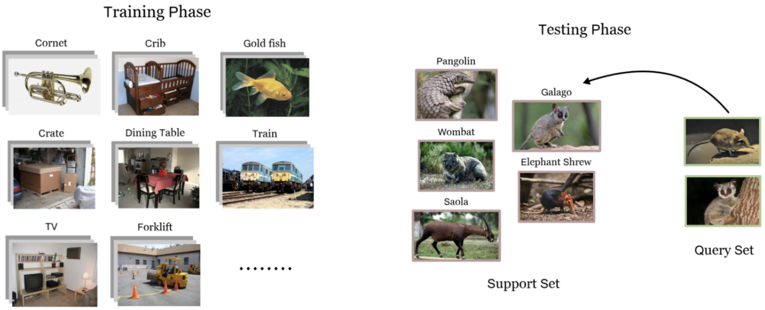
- N: number of classes.
- K: number of examples per class (small)

each classification task consists of a training set, also called **support set**, and a testing set, also called **query set**



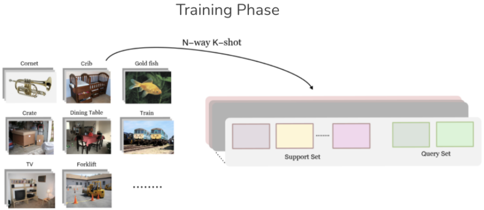
Few-shot learning

Few shot classification usually involves two stages, a training phase where we have access to the meta training set with a large number of classes, then a testing phase where we evaluate the learner on a set of tasks with novel classes not seen during training



Meta-Learning: initialize then adapt

Popular approach for few shot classification : the training phase matches the testing phase, where the learner in trained on a number of sampled tasks created from the meta training set.



Few-shot learning

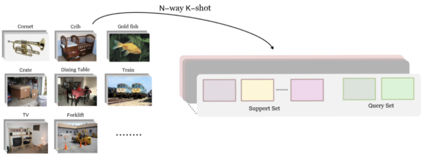
Meta-Learning: initialize then adapt

Training phase has matching conditions compared to the testing phase.

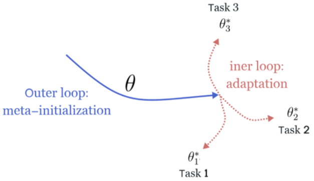
- The training set is transformed into many training tasks
- The learner is then trained on a distribution of such tasks (episodes)

Such methods share a common two loop structure, where the outer loop provided a meta initialization for the networks parameters, and from this meta initialization, the inner loop adapts these parameters for each task separately.

Training Phase

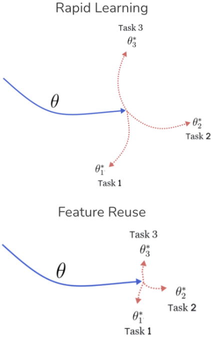


Model-Agnostic Meta-Learning



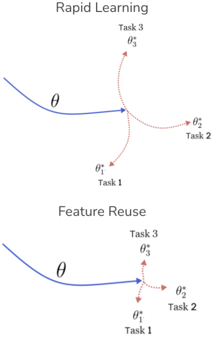
Few-shot learning

- Is this inner/outer loop necessary?
- How much of the effectiveness of such methods is contingent on the inner/outer loop structure?
- **Rapid Learning** : the representations change dramatically for each task and the inner loop plays an important role
- **Feature Reuse** : the outer loop gives rise to general purpose representations that require little adaption for each task

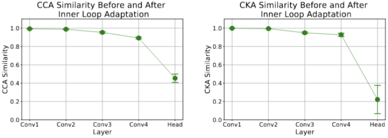


Few-shot learning

Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. Raghu et al. 2020



Similarity of the representation before and after adaptation

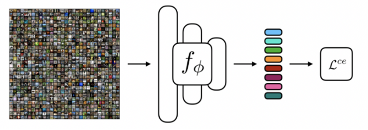


-> Feature reuse: Having general purpose representations is important for few-shot recognition

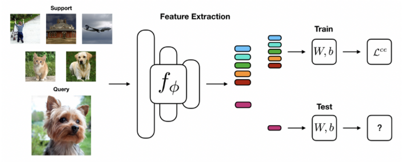
Transfer Learning: a simple and effective baseline

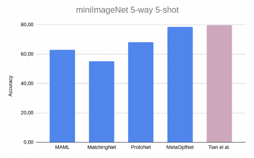
- At training time, train our model on the whole meta training set without episodic training.
- At test time, for each sampled task, use the pretrained and frozen feature extractor, then train a simple classifier

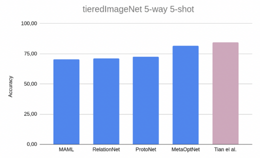
Training Phase



Testing Phase







Other data limitations

Other data limitations

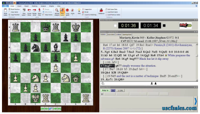
- Data bias - Fairness
- Data privacy
- Continuous world :
  - Incremental learning : input data is continuously used to extend the existing model's knowledge i.e. to further train the model. Goal : adapt to new data without forgetting its existing knowledge.
  - Continual and Lifelong learning : developing techniques and architectures enabling the models to learn sequentially without the need to re-train from scratch - Inspired from the human brain
- ...

Hybrid AI

Avant propos : un exemple simple

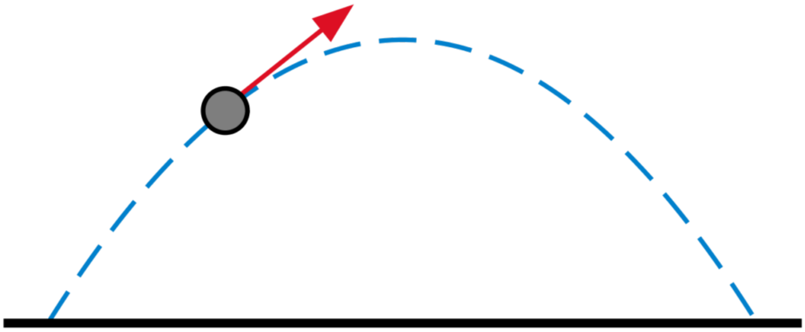
Jouer aux échecs

- Nécessite de connaître les règles du jeu
- **Approche symbolique ou basée connaissance**: modélisation explicite des règles du jeu avec un modèle à base d'états à partir duquel on peut décider si un coup est autorisé ou non.
- **Approche basée données** : apprentissage des règles du jeu à partir de données d'observation du jeu : possibilité de faire des coups illégeux ou de ne jamais atteindre certains coups légaux.

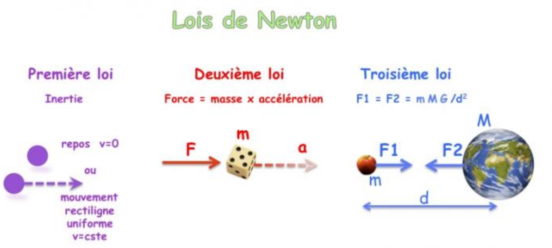


Un autre exemple simple

Prédire la trajectoire d'une balle !



Prédire la trajectoire d'une balle ! - Approche du physicien

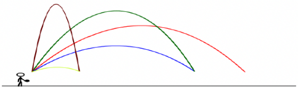


- Modélisation physique par une mise en équation du phénomène physique (e.g. EDP) et sa résolution numérique.
- Essais réels sur le système physique considéré.

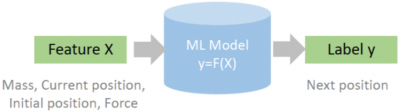
Prédire la trajectoire d'une balle ! - Approche du data scientist

Deux étapes

1. Lancer la balle autant de fois que de fois que possible: données d'observation du phénomène



2. Construire un modèle par apprentissage à partir des données d'observation



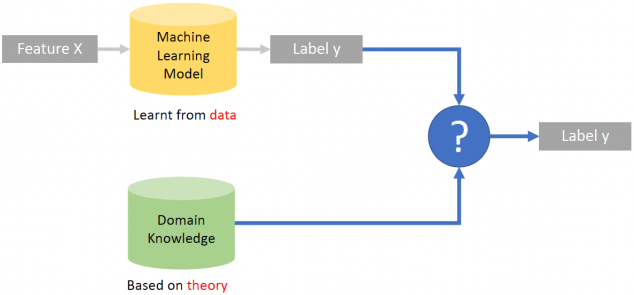
Limitations de l'approche données

- Les modèles pilotés uniquement par les données peuvent ne pas répondre à des contraintes telles que :
  - celles dictées par les lois naturelles,
  - ou données par des directives réglementaires ou de sécurité.

Trustworthy AI : pour une IA digne de confiance.

Approche hybride

Une solution : hybrider les modèles de connaissances (ici modèles physiques) avec l'IA (et plus particulièrement le machine learning)



Approche hybride

Lecture recommandée

Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems

Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Julius Pfrommer, Annika Pick, Rajkumar Ramamurthy, Michal Walczak, Jochen Garcke, Christian Bauckhage and Jannis Schuecker

Abstract—Despite its great success, machine learning can have its limits when dealing with insufficient training data. A potential solution is the additional integration of prior knowledge into the training process which leads to the notion of *informed machine learning*. In this paper, we present a structured overview of various approaches in this field. We provide a definition and propose a concept for informed machine learning which illustrates its building blocks and distinguishes it from conventional machine learning. We introduce a taxonomy that serves as a classification framework for informed machine learning approaches. It considers the source of knowledge, its representation, and its integration into the machine learning pipeline. Based on this taxonomy, we survey related research and describe how different knowledge representations such as algebraic equations, logic rules, or simulation results can be used in learning systems. This evaluation of numerous papers on the basis of our taxonomy uncovers key methods in the field of informed machine learning.

Index Terms—Machine Learning, Prior Knowledge, Expert Knowledge, Informed, Hybrid, Neuro-Symbolic, Survey, Taxonomy.

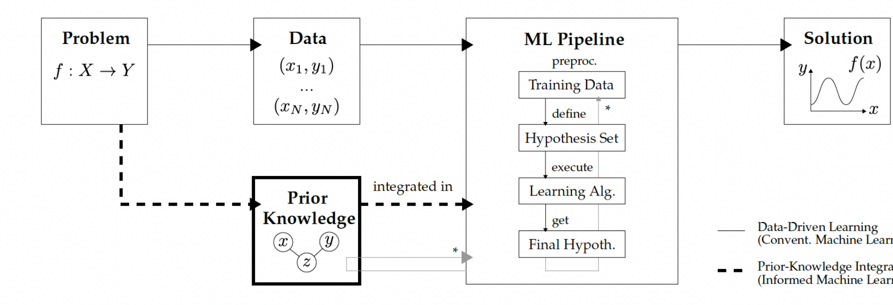
`https://arxiv.org/pdf/1903.12394.pdf`

Apprentissage informé

Définition [Rueden et al, 2021]
Apprentissage automatique à partir d'une source d'information hybride composée de **données** et **connaissances a priori**. Les connaissances proviennent d'une source indépendante, sont représentées formellement et sont intégrées de manière explicite dans le processus d'apprentissage

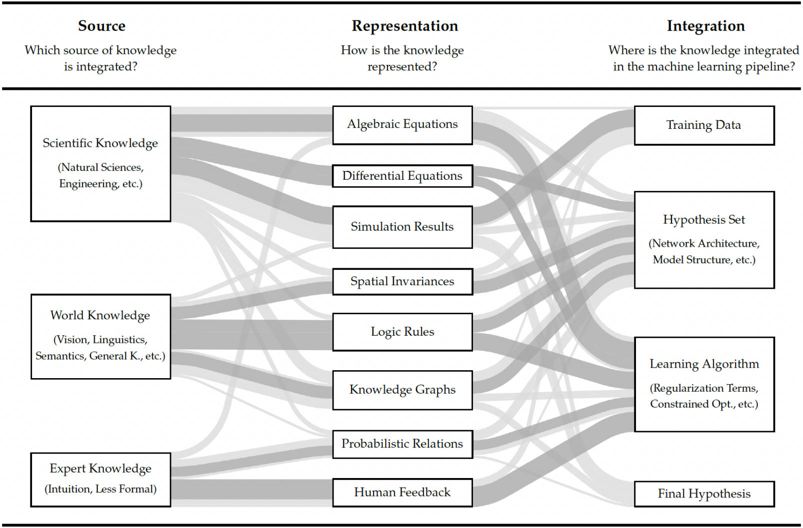
Hybrider apprentissage et connaissances

Principe [Rueden et al, 2021]



1. **Source** : quelle source de connaissances ?
2. **Representation** : Comment est représentée la connaissance ?
3. **Integration** : Où la connaissance est-elle intégrée dans la pipeline ?

Hybrider apprentissage et connaissances



[Rueden et al, 2021]

Physics-aware AI

Physics-Informed Machine Learning

Physics-informed Neural Networks (PINNs)

Des réseaux de neurones respectant les lois de la physique

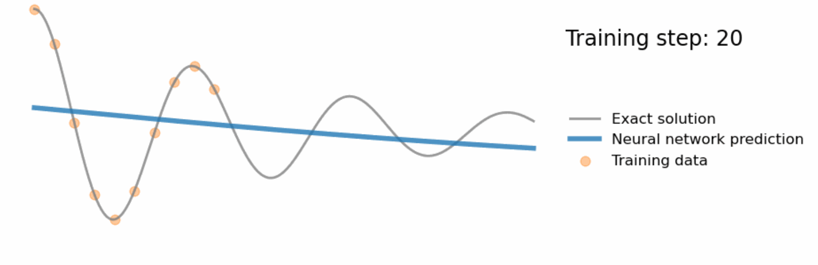
Motivation

- Faible efficacité des approches d'apprentissage pour résoudre des problèmes scientifiques et d'ingénierie complexes.
  - Les algorithmes ne prennent pas forcement en compte des principes physiques gouvernant les systèmes auxquels ils sont appliqués.
  - Prévoir toutes les conditions possibles est difficile.
- **Idée** : rendre les algorithmes d'apprentissage plus robustes et performants en intégrant les lois de la physiques et les connaissances scientifiques.

Physics-informed Neural Networks (PINNs)

Exemple

Données expérimentales d'observation d'un phénomène physique inconnu (training data). On veut prédire précisément les nouvelles mesures.

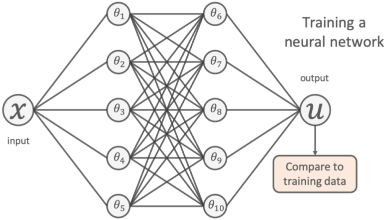


Source : Ben Moseley (`https://benmoseley.blog/my-research/so-what-is-a-physics-informed-neural-network/`)

Physics-informed Neural Networks (PINNs)

Exemple

Approche classique : par apprentissage avec un réseau de neurones.



Minimisation de l'erreur aux moindres carrés entre la prédiction et sa vérité-terrain.

$$\min \frac{1}{N} \sum_{i}^{N} (u_{NN}(x_i; \theta) - u_{\mathrm{true}}(x_i))^2$$

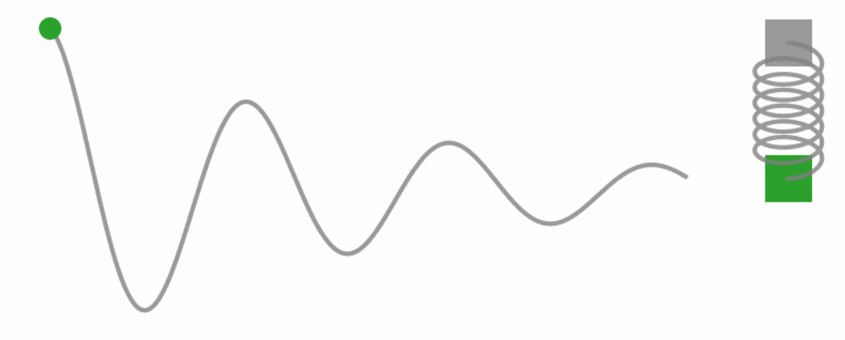
Problème : Généralisation en dehors des données d'apprentissage

Source : Ben Moseley (`https://benmoseley.blog/my-research/so-what-is-a-physics-informed-neural-network/`)

Physics-informed Neural Networks (PINNs)

Exemple

Supposons que l'on a des connaissances sur la physique du système observé : les points sont en fait des mesures de la position d'un oscillateur harmonique amorti.



$$m\frac{d^2u}{dx^2} + \mu\frac{du}{dx} + ku = 0$$

avec m, la masse de l'oscillateur, μ, le coefficient de friction et k est la constante du ressort.

Source : Ben Moseley (`https://benmoseley.blog/my-research/so-what-is-a-physics-informed-neural-network/`)

Physics-informed Neural Networks (PINNs)

Exemple

Idée : ajouter les équations différentielles connues directement dans la fonction de perte utilisée pour l'entrainement. Ajout d'une terme dans la fonction de perte qui vise à garantir que la solution apprise par le réseau est cohérente avec la physique connue.

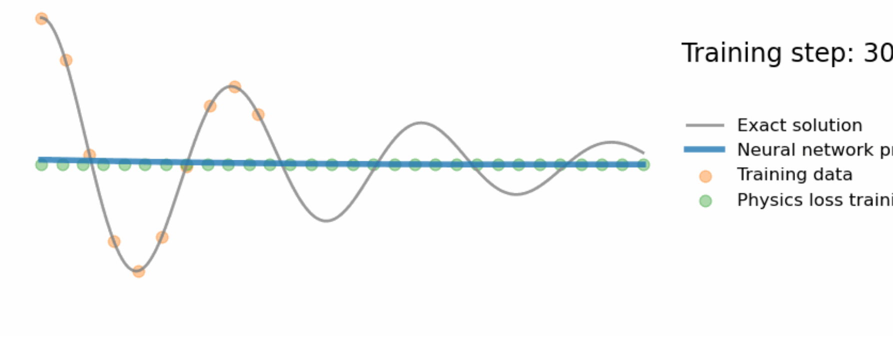
$$\min \frac{1}{N}\sum_{i}^{N}(u_{NN}(x_i;\theta) - u_{\mathrm{true}}(x_i))^2$$

$$+\frac{1}{M}([m\frac{d^2}{dx^2} + \mu\frac{d}{dx} + k]u_{NN}(x_j;\theta))^2$$

Source : Ben Moseley (`https://benmoseley.blog/my-research/so-what-is-a-physics-informed-neural-network/`)

Physics-informed Neural Networks (PINNs)

Exemple



Source : Ben Moseley (`https://benmoseley.blog/my-research/so-what-is-a-physics-informed-neural-network/`)

Multimodal AI

Multimodal AI

Goal

Build models that can process and relate information from **multiple modalities**.
Learning from multimodal sources offers the possibility of capturing correspondences between modalities and gaining an in-depth understanding of natural phenomena

Core challenges

- **Representation**: representing data in an efficient and meaningful way from multiple modalities that often contain both complementary and redundant information.
- **Translation** :challenge of translating data from one modality to another.
- **Alignment** : task of establishing correspondence between data from two or more modalities of the same event.
- **Fusion** : joining information from two or more modalities to perform a given task.
- **Co-learning** : challenge of transferring knowledge between modalities
- **Interpret and explain**

AI and explainability

Does an AI have to explain itself?

Black boxes pose the problem of bias in the data.

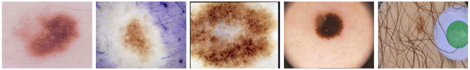
Skin images

- The ISIC database is a database of annotated dermoscopic images.
- Use in different challenges.
- Deep neural networks (AUC=71%) have better results than dermatologists (AUC=67%)

(Bissoto el al, 2019) (De)Constructing Bias on Skin Lesion Datasets (https://openaccess.thecvf.com/content_CVPRW_2019/papers/ISIC/Bissoto_DeConstructing_Bias_on_Skin_Lesion_Datasets_CVPRW_2019_paper.pdf)

Does an AI have to explain itself?

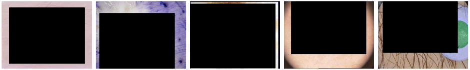
Black boxes pose the problem of bias in the data.



(a) Traditional images

(b) Only Skin images

(c) Bbox images



(d) Bbox70 images

Does an AI have to explain itself?

Black boxes pose the problem of bias in the data.

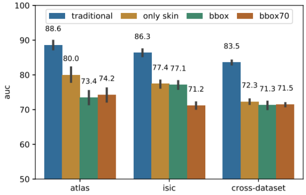


Figure 3: Models' performance over the disturbed datasets. We first remove all the pixel colors inside the lesion (*only skin*), proceeding to remove border information (*bbox*), and finally, removing the size (diameter) of the lesion (*bbox70*). Surprisingly, even when we destruct all clinical-meaningful information, the network finds a way to learn to classify skin lesion images much better than chance.

The Need of explanations

- AI systems are often involved in **decision-making tasks**, a task which, in its nature, need explanations.
  - More than the output(s) of a prediction: provide elements of evidence, explanatory elements.
- **Legal aspects**: RGPD (right to explanation)
- **Human agency and oversight** : Make the decision transparent, understandable and explainable: to allow interaction and communication with users.
- **For trust and acceptation**.

Explainability as part of a growing Global AI Policy and Regulation

- UNESCO : Elaboration of a Recommendation on the ethics of artificial intelligence[2]
- European Commission : Ethics guidelines for trustworthy AI[3]
- European GDPR: Article 22 empowers individuals with the **right to demand an explanation** of how an automated system made a decision that affects them.
- California Consumer Privacy Act (CCPA) : Requires companies to rethink their approach to **capturing, storing, and sharing personal data** to align with the new requirements by January 1, 2020.
- Algorithmic Accountability Act 2019[4] : Requires companies to provide an **assessment of the risks** posed by the automated decision system to the privacy or security and the risks that contribute to inaccurate, unfair, biased, or discriminatory decisions impacting consumers.
- ...

[2] `https://unesdoc.unesco.org/ark:/48223/pf0000373434`

[3] `https://://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai`

[4] `https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info`