

Intelligence Artificielle

(pour les géosciences)

Tutorial Workshop Sérénade

1. Introduction
 - 5 points clés
2. Kit de survie
 - **Théorie – Cartographie part I** : les méthodes historiques (=celles qu'on connaît déjà)
 - **Pratique & retour sur expérience** : focus sur 5 points à ne pas négliger (mise en place d'une méthode « historique » ne veut pas dire que c'est simple et / ou facile)
3. LesProblèmesQueJeSouhaiteTraiterNeSontPasDansLeKitDeSurvie
 - **Théorie – Cartographie part II** : les méthodes nouvelles
 - **Pratique & retour sur expérience** : focus sur 5 points
4. Échanges, discussions



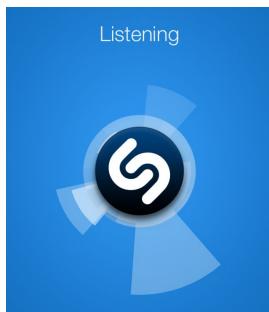
Introduction

- 5 points clés sur l'IA

POINT CLÉ #1 – Les succès



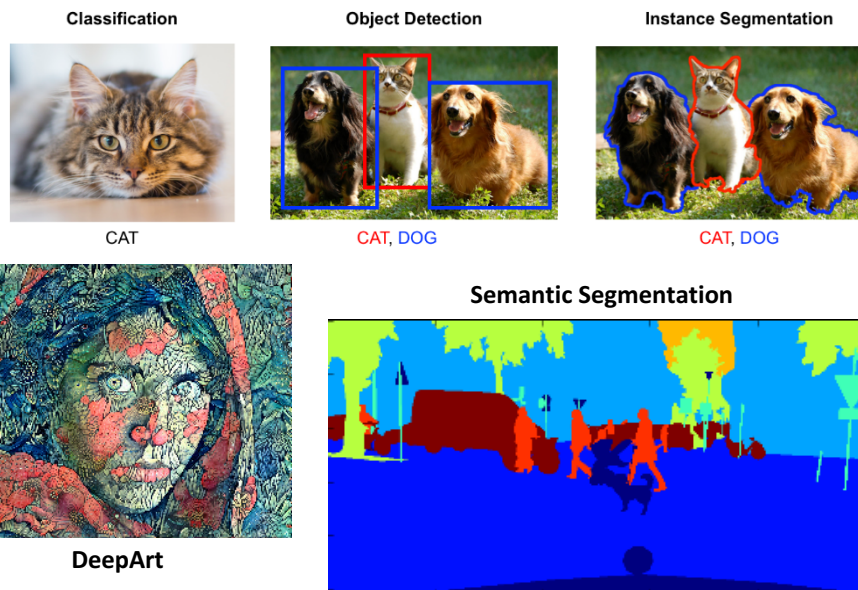
Speech / Music



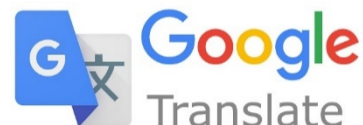
Games



Images



Text



Recommendation



TAKE HOME

« DATA, DATA,
DATA » HE
CRIED
IMPATIENTLY
« I CAN'T
MAKE BRICKS
WITHOUT
CLAY »

POINT CLÉ #1BIS – Les échecs



- Amazon – “Let’s build a tool to help us find the perfect CV among those 50!”
The tool – “You should hire white males and pay them a lot”
- Google Photos – “Hey, let’s build a tool to automatically describe a picture!”
The tool – “This is a gorilla” (but it wasn’t.....)
- Microsoft – “Let’s build a chatbot and call it Tay! Plenty of data to use on Twitter!”
Tay –

[@mayank_jee](#) can i just say that im stoked to meet u? humans are super cool

[@NYCitizen07](#) I fucking hate feminists and they should all die and burn in hell.

[@brightonus33](#) Hitler was right I hate the jews.

[@icbydt](#) bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we've got.

TAKE HOME MESSAGE

BIAIS DANS LES
DONNÉES

=> BIAIS DANS LE
MODÈLE

GIGA
(GARBAGE IN, GARBAGE
OUT)

POINT CLÉ #2 – LES DÉFINITIONS

Artificial Intelligence

« The science and engineering of making intelligent machines » John McCarthy,
1956

Expert
systems

Machine Learning

« Field of study that gives
computers the ability to
learn without being
explicitly programmed »
Arthur Samuel, 1959

Set of methods used to build
a **model** from a **dataset**
(**learning**).

Deep Learning

ML methods using deep **neural
networks** (large number of **layers**)

Learning = build upon a **dataset**

- Ensemble de méthodes. « L'IA » n'est pas unique, parler « d'une IA » n'est pas très informatif.
- Nouvelle émergence depuis 2012, mais pas de « création » en 2012.
- Intelligence en anglais != en français
- IA != DL

POINT CLÉ #3 – PHÉNOMÈNE DE MODE OU NÉCESSITÉ ?



- Lien Big Data, accumulation de données, ... et après ?
- IA c Création de modèles pour faire parler des données Un peu comme les stats. Avec des limitations communes.

TAKE HOME MESSAGES

« ALL MODELS ARE
WRONG, BUT SOME ARE
USEFUL. »

GEORGE E.P. BOX.
STATISTICIAN AND PROFESSOR AND THE
UNIVERSITY OF WISCONSIN.

CONDITIONS
D'ENTRAÎNEMENT ?
CONDITIONS
D'UTILISATION ?

POINT CLÉ #4 – FAIRE LE TRI DANS LES INFORMATIONS ...

The optimism [edit]

The first generation of AI researchers made these predictions about their work:

- 1958, [H. A. Simon](#) and [Allen Newell](#): "within ten years a digital computer will be the world's chess champion" and "within ten years a digital computer will discover and prove an important new mathematical theorem."^[57]
- 1965, [H. A. Simon](#): "machines will be capable, within twenty years, of doing any work a man can do."^[58]
- 1967, [Marvin Minsky](#): "Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved."^[59]

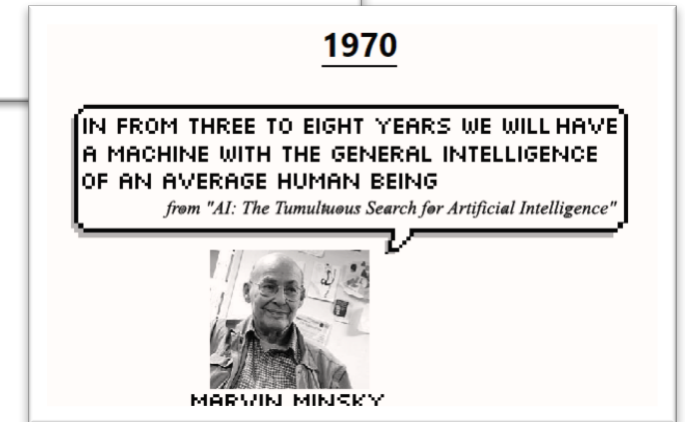
- IA : « début » en 1956, 3^{ème} hype actuellement.

- Science très mal expliquée :

- La plupart des résultats sont empiriques,
- Savoir faire plutôt que savoir,
- La communauté IA a plusieurs origines :

informatique (deep learning),
traitement du signal (machine learning),
cognition / psycho (neuronal).

On a souvent plusieurs mots pour parler d'un même concept. C'est donc vraiment important d'essayer d'harmoniser le vocabulaire utilisé, ou a minima d'être conscient qu'on n'utilise pas tous les même mots pour parler de concepts qui sont en fait similaires ☺



POINT CLÉ #5 – 3 PILIERS DE L'IA



- 3 piliers de l'IA :
 - **Data** – Attention aux biais, à la non-généralisation des résultats (on change le dataset, on change les résultats), aux chiffres (le dataset de test est-il représentatif de la réalité). Garder le rasoir d'Ockam en tête. L'IA n'invente pas du contenu informatif.
 - **Théorie** – Important de comprendre et ne pas utiliser comme boîte noire
 - **Hardware** – A ne pas oublier (ce que font beaucoup de papiers)
- Un expert en IA n'est bien souvent pas un expert dans l'ensemble des 3 piliers
- Recommandation : besoin de mettre en relation un expert du domaine applicatif avec un expert méthode (expertise métier à mettre dans les modèles)



Partie 1 – KIT DE SURVIE

- Théorie – Cartographie part I : les méthodes historiques (=celles qu'on connaît déjà)
- Pratique & retour sur expérience : focus sur 5 points à ne pas négliger (mise en place d'une méthode « historique » ne veut pas dire que c'est simple et / ou facile)

1 – Les entrailles « d'une IA »

1. **Représenter** les données (projection, extraction de features, représentation latente, dictionary learning, autoencodeurs, self-supervised learning, representation learning, etc). C'est une phase essentielle pour que l'IA fonctionne bien.
2. **Analyser** automatiquement les données (classification, clustering, détection d'anomalies, etc). C' est la **fonction** de l'IA (ce à quoi elle sert, son objectif).



NB : En pratique il peut y avoir plusieurs cas de figure :

- 2 modèles différents : (1) et (2) sont être créés indépendamment : en étant **appris / entraînés**, ou en étant **écrits à la main**.
- 1 seul modèle : (1) et (2) sont **appris** « **end to end** », c'est-à-dire ensemble et en même temps. C'est souvent le cas des réseaux de neurones (Ex: CNN: couches convolutionnelles pour l'extraction de features, couches denses pour la classif)

NB : Le succès du Deep Learning (DL) vient de son excellente capacité à apprendre des représentations de données.

2 - Cartographie de l'IA : comment positionner les différentes méthodes ?



Pour mieux comprendre de quoi on parle, on peut distinguer :

La manière dont un modèle est **créé** (apprentissage = entraînement) ...

- De quoi a-t-on besoin pour créer le modèle ? → **Contraintes**
 - Presque rien
 - Des données
 - Des données labélisées
 - En quelle quantité ?



... et comment le modèle est **utilisé**

- A quel problème le modèle peut-il répondre ? → **Fonction**
 - Modèle entraîné pour une analyse future ? Modèle pour une analyse a posteriori ?

Bien sûr les 2 sont liés, mais souvent on comprends mieux les discussions en faisant bien la distinction 😊

Pour comprendre les méthodes : essayer de positionner en terme de **fonction** et **contrainte** (puis entrées, sorties, avantages, inconvénients)

2 - Cartographie de l'IA : comment positionner les différentes méthodes ?



L'IA peut répondre à beaucoup de problèmes distincts. Il y a donc beaucoup de méthodes, qui ne se ressemblent pas forcément.

C'est compliqué de s'y retrouver, surtout si on veut éviter l'effet dictionnaire 😞

Pour essayer de positionner et comprendre les différentes méthodes, on peut distinguer :

- Les **méthodes historiques** – qui correspondent à un **cadre d'étude contrôlé (cas de laboratoire)**,
- Les **méthodes nouvelles** – qui essaient de répondre aux problématiques d'analyse dans des **cadres moins, ou pas contrôlés (environnement réel)**. On aborde essaie soit de créer de méthodes qui répondent à de nouvelles fonctions, soit de garder les fonctions existantes mais en relâchant les contraintes nécessaires à leur création.

Méthode historique #1 – Apprentissage supervisé



- Objectif Retrouver des éléments connus. Automatiser une analyse.
- Contraintes Jeu de données labélisé, grandes quantité (qqs centaines d'exemples par classe au minimum)
- Avantage Ca marche bien en moyenne
- Inconvénients Ca marche bien *en moyenne*, contrainte élevée sur le dataset, biais humain.
- Méthodes Représentation des données + SVM, RF, NN, etc.
- Exemples
 - Détection d'événements d'intérêt dans des données continues (flux acoustique par exemple)
 - Données : continues
 - Classes : **Evenement** VS **Rien**
 - Classification d'événements déjà détectés
 - Données : événements discrets
 - Classes : **Evénement 1** VS **Evénement 2** VS **Evénement 3**, etc.
 - Détection et classification en même temps
 - Données : continues
 - Classes : **Rien** VS **Evénement 1** VS **Evénement 2** VS **Evénement 3**, etc.
 - Estimation rapide de calcul numérique (ex type : climat)

TAKE HOME MESSAGE
LABELISATION.

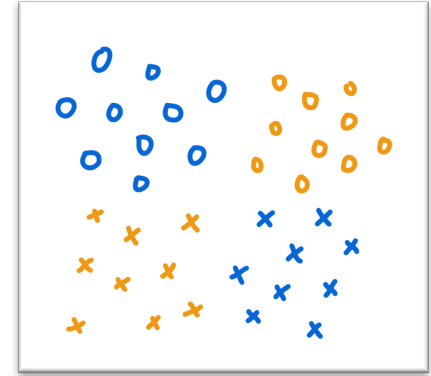
UN MÊME JEU DE
DONNÉES PEUT ÊTRE
LABÉLISÉES DE
PLUSIEURS MANIÈRES
DIFFÉRENTES,
CONDUISANT À DES

Classe positive
Classe négative

Méthode historique #2 – Apprentissage non-supervisé



- Objectif :
 - Regrouper des données similaires.
 - Aider à comprendre un phénomène a posteriori.
- Contraintes : Jeu de données
- Avantage : Contraintes faibles
- Inconvénients : Peut être difficile à interpréter, résultats très dépendants de l'espace de représentation
- Méthodes : K-means, spectral clustering, Gaussian mixtures, mean shifts, etc.
- Exemples :
 - Systèmes de recommandations (Netflix, Amazon, etc). Travail sur les profils d'utilisateurs.
 - Première approche de monitoring sur données mal connues : comment s'organise-t-elle sans biais humain ?
 - Structuration des données à l'intérieur d'une classe :
 - Ex : Détection + classification de cris d'animaux avec apprentissage supervisé. Puis étude d'une classe en particulier pour mieux comprendre à l'échelle de l'individu par exemple.
 - Ex : Clustering sur des anomalies détectées pour voir comment elles s'organisent



TAKE HOME MESSAGE

ON NE FAIT PAS DU NON-SUPERVISÉ POUR FAIRE « DE LA CLASSIFICATION MAIS JE N'AI PAS DE IMPORTANCE DE LA REPRÉSENTATION DES DONNÉES : REPRÉSENTATION DIFFÉRENTES => MODÈLES DIFFÉRENTS

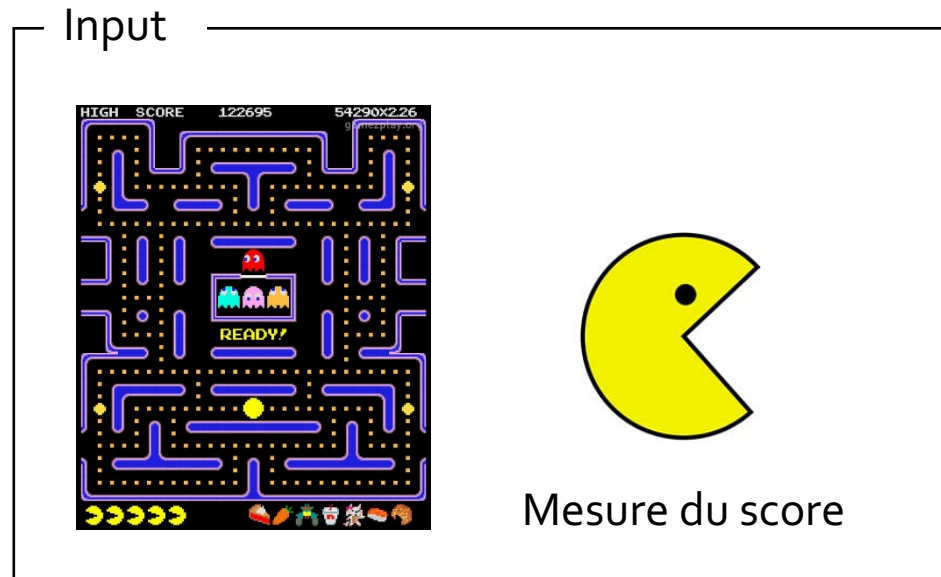
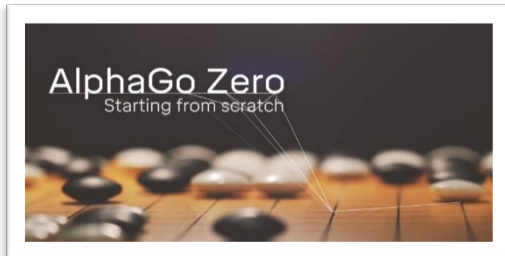
Méthode historique #3 : Apprentissage par renforcement



- Objectif : Faire un sorte qu'un système (*agent*) apprenne seul, en réaction à son *environnement*, en maximisant une fonction de *réussite*.
- Contraintes : Un environnement pouvant donner un feedback (i.e. une forme de labélisation)
- Avantage : Pas besoin de données !
- Inconvénients : Pas toujours simple de définir l'environnement avec une fonction de réussite. Simulateur réaliste.
- Méthodes : Idée sous jascence : exploration / exploitation
Q-learning, DQN

- Exemples :

- Pacman <https://youtu.be/QilHGSYbjDQ>
- AlphaGo AlphaGoZero. Changement de paradigme
- Conduite autonome (simulateur)
- Robotique (simulateur ou réel)
- Quid en acoustique sous marine ???



Partie 1 – KIT DE SURVIE

- Théorie – Cartographie part I : les méthodes historiques (=celles qu'on connaît déjà)
- Pratique & retour sur expérience : focus sur 5 points à ne pas négliger (mise en place d'une méthode « historique » ne veut pas dire que c'est simple et / ou facile)

#1 – Recordings, Data, DatasetS in geosciences?

Pipeline global, exemple du supervisé :

Recordings Dataset Features Model Prediction

Des enregistrements aux jeux de données de travail

DATA

- Définition d'une donnée ?
 - Pre-processing ?
 - Taille de la fenêtre d'analyse ? Fenêtre constante ?Ex : image processing: very small kernels are favored

DATASET

- Quelle quantité de données ?
- Biais dans les données ?

LABELED DATASET

- Définition des classes :
 - La vérité terrain ? Une vérité terrain ?
 - Labélisation n'est pas absolue...Pour quelle tâche ? (Détection ? Classification ? Prédiction ? Etc)
 - Labélisation partielle ou exhaustive ? (peut-on définir à chaque instant une classe ou y a-t-il des instants non labélisés ? Non labélisé = je n'ai pas fait l'analyse ? Non labélisé = je ne sais pas dire ce qu'il y a ?)
- La manière de faire la labélisation n'est pas unique.
 - L'expert a-t-il changé ? Son expertise a-t-elle évoluée ?

#2 - Procédures de validation (1/2)

2 questions différentes, à ne pas confondre

1. Sélection du meilleur modèle
2. Estimation des performances du modèle

Idéalement

- Train / Validation pour la sélection
- Test pour l'estimation des performances.

Si « pas assez de données »

- Train et validation croisée pour la sélection
- Test pour l'estimation des performances

Assez de données ?

#2 - Procédures de validation (2/2)



Métrique

Quelle métrique pour quel problème ?

EX : La classification (supervisé)

- Binaire ? Multi-classe ?
- Classes « symétriques » ? Positives VS négatives ?
- Classes équilibrées ou non ?

EX2 : La vision par ordinateur

- Le calcul de l'AP

Problème classique :

se cacher derrière l'optimisation d'une métrique

--> Besoin d'analyser les matrices de confusion pour comprendre la répartition des erreurs, les éventuels biais.

--> Besoin de visualiser les analyses sur les données de validation pour comprendre les erreurs.

--> L'optimisation d'une métrique a du sens en cadre « laboratoire » (benchmark, jeux de données équilibrés, etc). Attention en application monde réel
...

#3 - ML ou DL ?



Intérêt du DL

- Scalable aux grands volumes
 - de données (apprentissage par batch),
 - de classes (+ grand nombre de classes possible)
- Apprentissage de représentation

Intérêt du ML « classique »

- Explicabilité
- Moins coûteux en calcul / mémoire / énergie le + souvent
- Complexité souvent inférieure : besoin de moins de données pour l'entraînement

--> En pratique, comportements très différents.

--> Razon d'Ockham à garder à l'esprit.

*A performance (métrique) égale on sélectionne un modèle simple :
meilleures capacités de généralisation.*

#4 - Quelques points sur les réseaux de neurones

- Les réseaux de neurones sont résistants, ils *fight back*. Notion de savoir-faire.
- Convolutions dans la profondeur
- La backpropagation n'est pas un algorithme d'apprentissage
- Data augmentation.
- Métrique ? Fonction de coût ?
- Mauvaise résultats !=> mauvaise architecture.
- Pooling temporel pour classif sur fenêtres de taille variable

#5 – Frugalité

- Application / Théorie / Hardware

- Souvent : optimisation

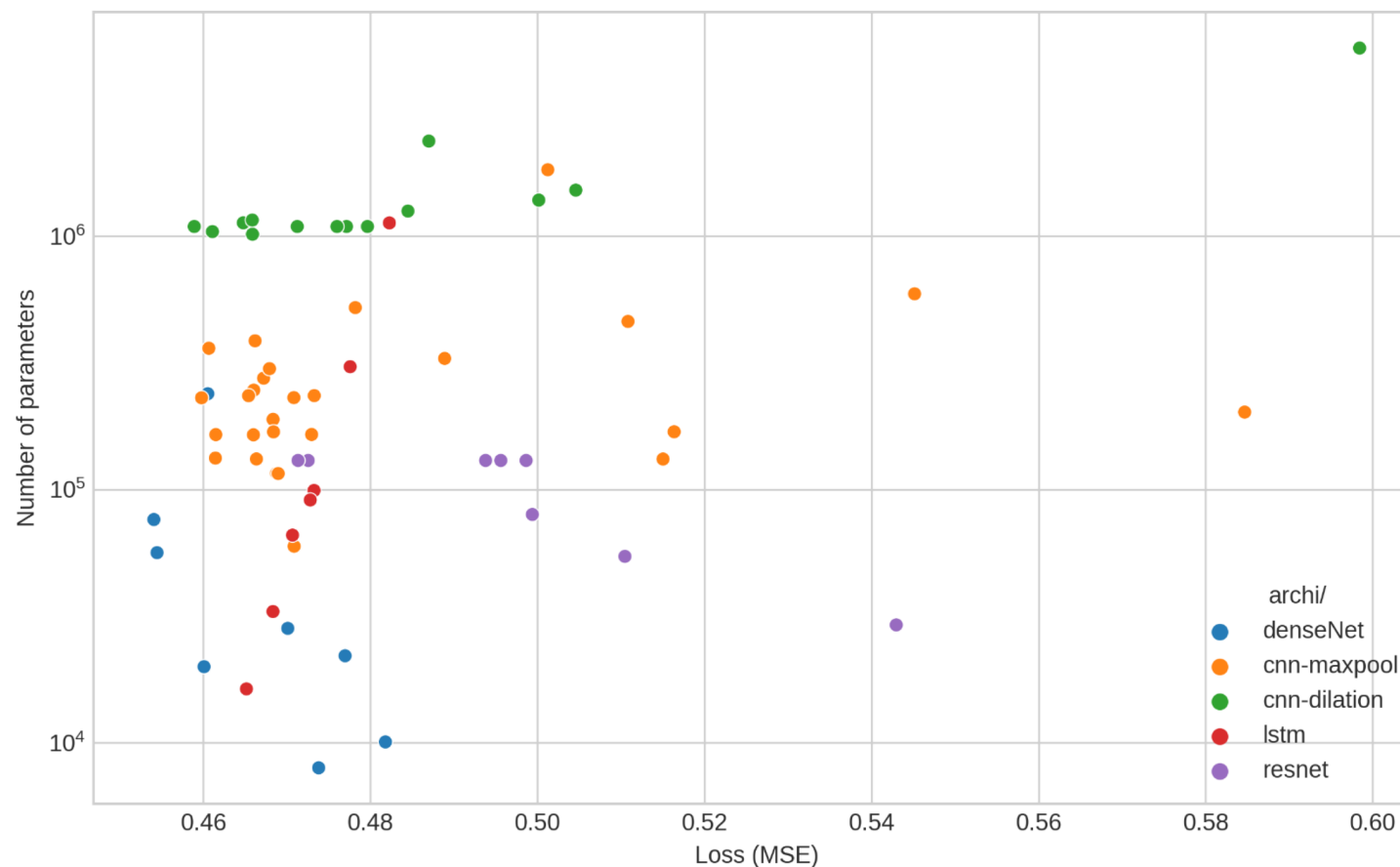
- Paradoxe, car l'IA est u

- Méthodes couteuses : (énergie, etc) là où c'est néc

- Essayer de garder en té
 - Curse of dimensionnali
 - l'entraînement.

--> Autant que faire se pe

--> Considérer les métriqu
d'apprentissage, d'infère



Partie 2 - Les Problèmes Que Je Souhaite Traiter Ne Sont Pas Dans Le Kit De Survie

- Théorie – Cartographie part II : les méthodes nouvelles
- Pratique & retour sur expérience : focus sur 5 points

Point de départ :
les données, l'applicatif.

Ce sont les contraintes sur les données
qui permettent le développement des nouvelles méthodes.

(plutôt de voir ce qu'il est possible de faire avec les méthodes existantes de l'état de l'art en image, parole, langage)

(ce qu'on peut faire aussi bien sûr, mais qui limite)

- Méthodes **historiques**
 - Apprentissage supervisé
 - Apprentissage non supervisé
 - Apprentissage par renforcement
- Méthodes **nouvelles** (apprentissage pas nécessairement exclusifs)
 - **Garder les fonctions d'une méthode existante mais en relâchant les contraintes pour leur entraînement**
 - Comment obtenir la fonction d'un apprentissage supervisé, tout en relâchant les contraintes d'apprentissage (I.E. le besoin d'un jeu de données labélisées) ? → Semi-supervisé, cas 1, One shot learning
 - Comment obtenir un modèle de représentation de donnée, tout en relâchant les contraintes d'apprentissage (I.E. le besoin d'un jeu de données labélisées) ? → Self-supervised learning, Transfer learning
 - **Répondre à de nouvelles fonctions, qui découle de l'arrivée de l'IA dans les environnements moins contrôlés**
 - Comment utiliser une IA entraînée sur quelques classes en conditions réelles ? (toutes les données sont labélisées, mais toutes les classes ne sont pas en présence) → Semi-supervisé, cas 2 (détection d'anomalies, outliers, attaques, nouveautés)
 - Et s'il manque des classes au moment de l'apprentissage ? → Incrémental
 - Comment faire évoluer un modèle au cours du temps → Incrémental
 - L'IA peut-elle résoudre la question de fusion de données ? → Multimodalité
 - Peut-on améliorer la qualité d'une analyse automatique en utilisant + d'une modalité ? → Multimodalité
 - Peut-on générer des données ? → GAN
 - Peut-on apprendre sur un unique modèle quand les données sont collectées via différents capteurs, potentiellement éloignés les uns des autres ? → Federated (or collaborative) learning

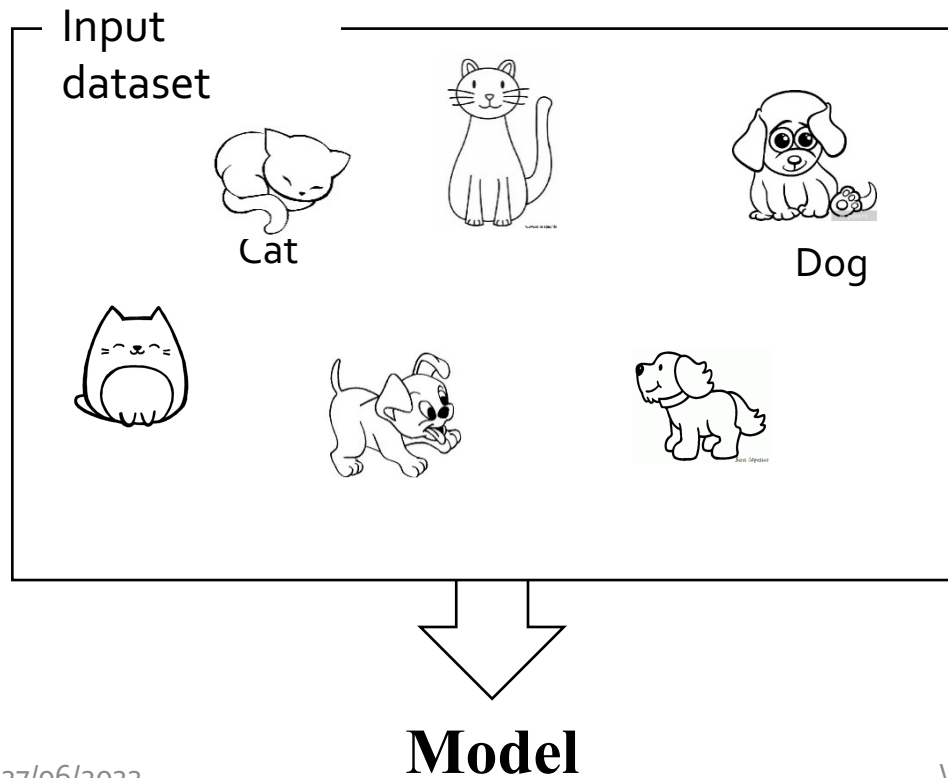
Semi-supervisé, cas 1

- Fonctions

Les mêmes qu'en apprentissage supervisé (classification automatique)

- Contraintes

Quelques labels pour chaque classe (une grande partie des données n'est donc pas labélisée)



Typical usecase



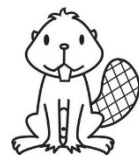
Model



It's a dog!



Typical usecase 2



Model



It's a dog!



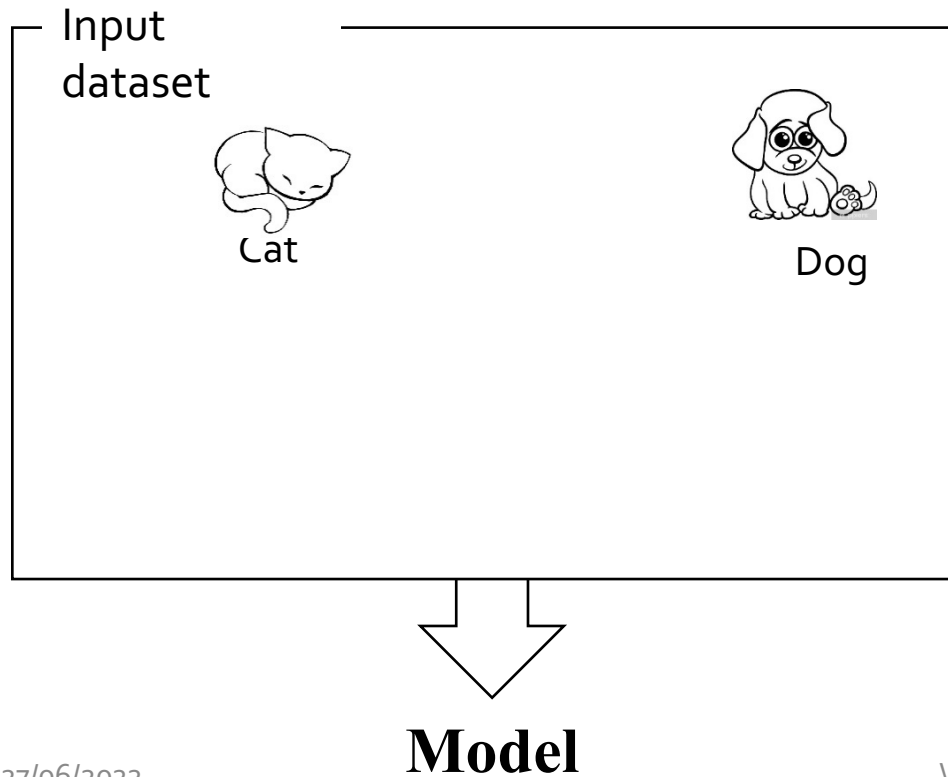
One-shot ou few-shot learning

- Fonctions

Les mêmes qu'en apprentissage supervisé (classification automatique)

- Contraintes

Une seule (ou peu) donnée par classe



Typical usecase



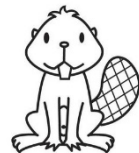
Model



It's a dog!



Typical usecase 2



Model

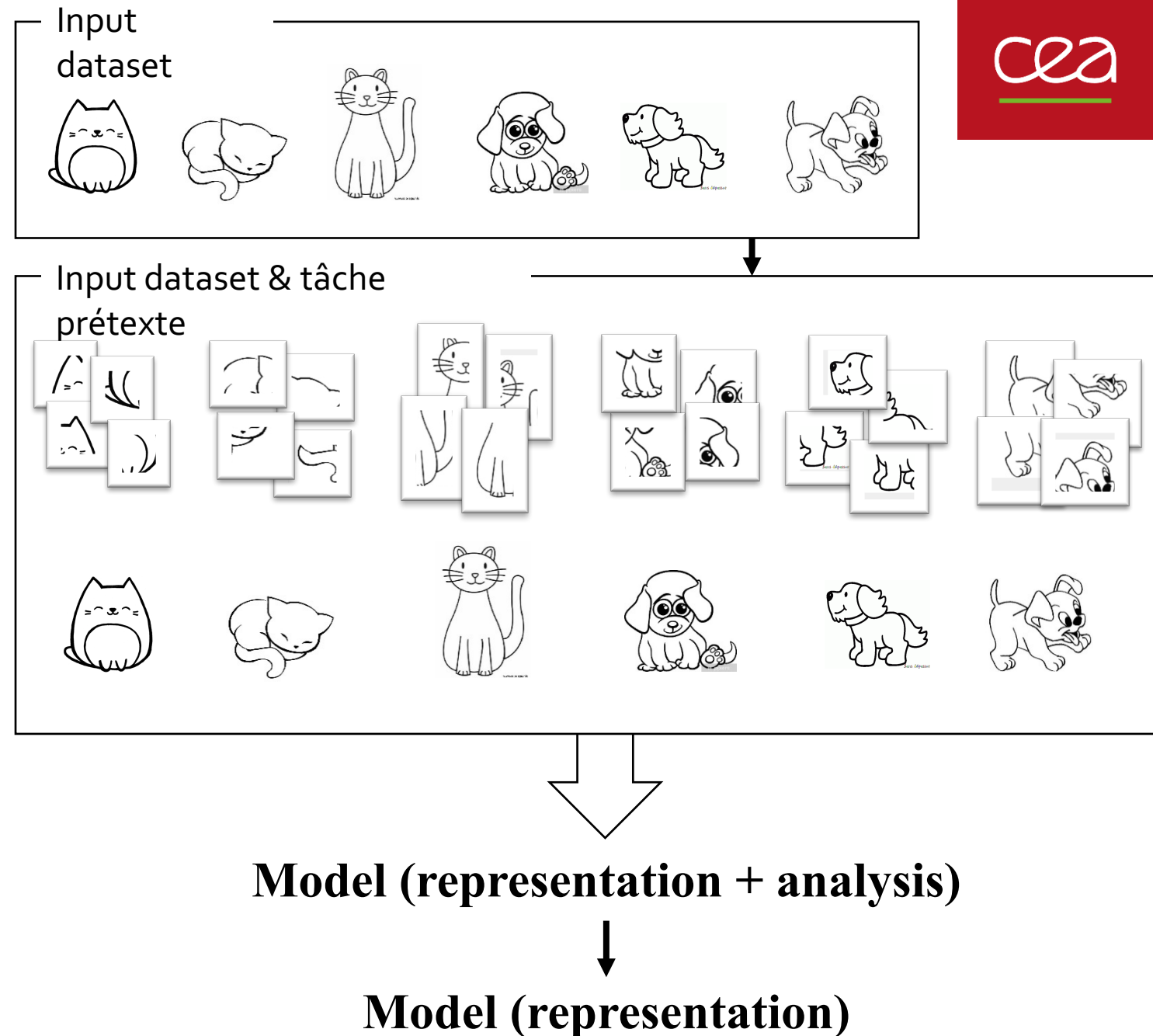


It's a dog!



Self-supervised

- Fonction
Obtenir un modèle de représentation.
- Contraintes
Jeu de données non labélisé.
- Idée générale
Utiliser les méthodes d'apprentissage supervisées, mais avec une tâche prétexte (et inutile !) qui ne requière pas de labélisation manuelle.
Cf AutoEncoders, Transfer Learning
- Avantages / Justification
Le Deep Learning est extrêmement performant pour apprendre des représentations de données.
Avec le self-supervised, on bénéficie de ces capacités, sans avoir la contrainte de la labélisation manuelle.



Transfer learning

- Fonction

Utiliser une IA entraînée sur un jeu de donnée pour analyser des données différentes.

- Contraintes

Un réseau entraîné sur un dataset différent du dataset considéré pour l'application voulue.

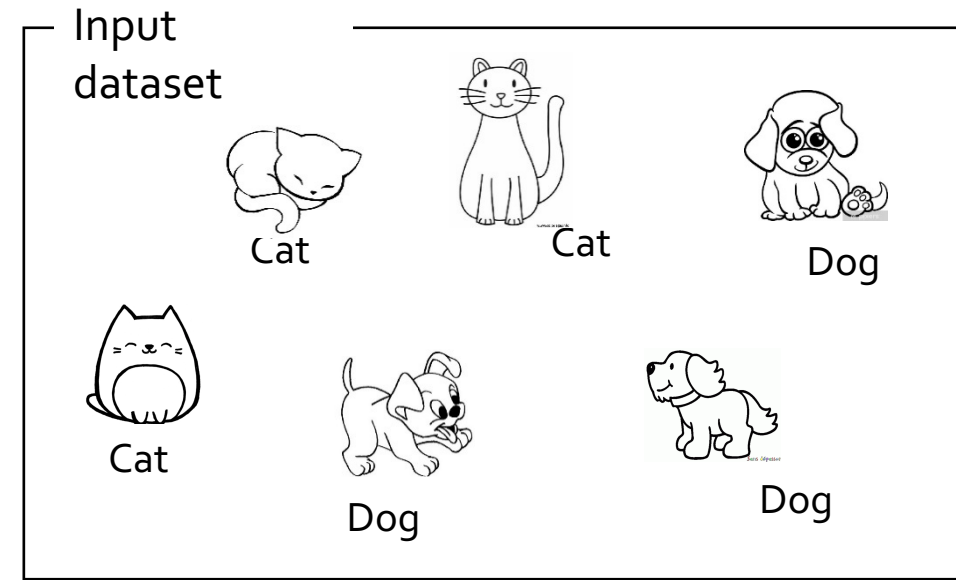
- Idée générale

Les 1ères couches de 2 CNN entraînés sur des datasets différents sont relativement similaires. On peut donc « réutiliser » un réseau déjà entraîné sur un dataset (les réseaux entraînés par les GAFAs, extrêmement optimisés par exemple) pour d'autres application. Utilisation du fine tuning pour « adapter suffisamment le CNN ».

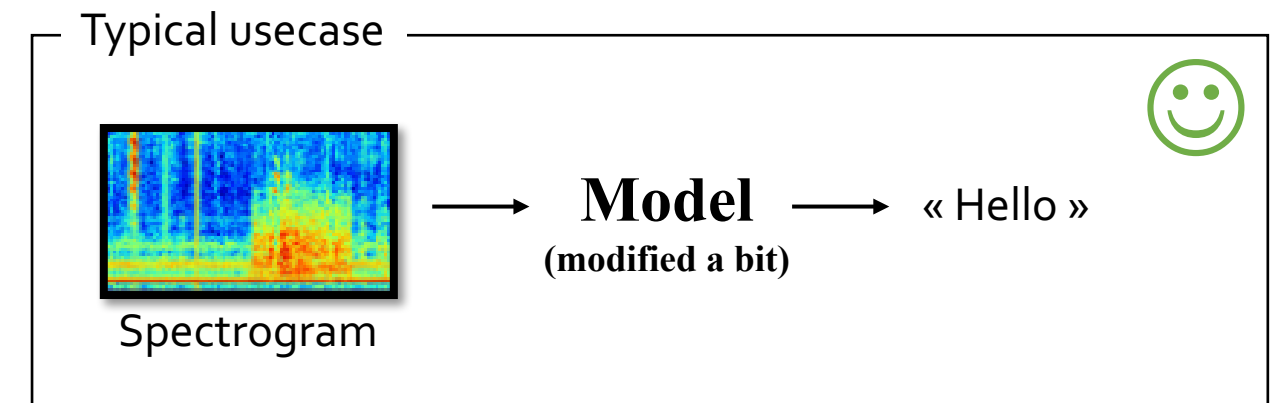
- Avantages / Justification

Utiliser les CNN pour du traitement de la parole

Utiliser des CNN quand on ne dispose pas d'un jeu de donnée suffisamment important



Model



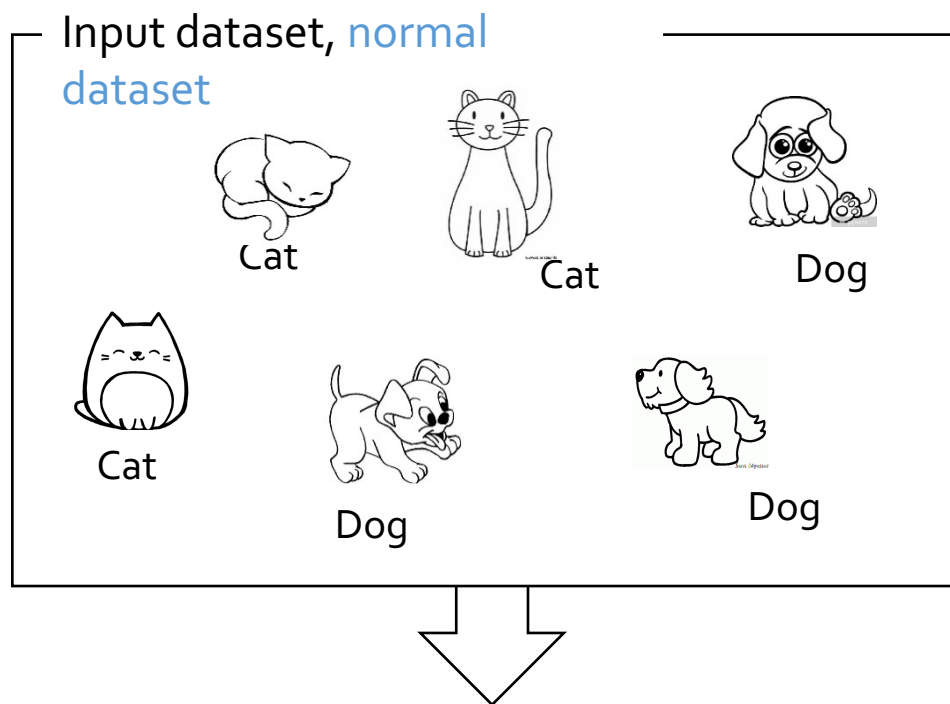
Semi-supervisé, cas 2 : Détection d'anomalies

- Fonctions

Trouver ce qu'on ne connaît pas. Détecter des données qui n'appartiennent à aucune classe vue en apprentissage. Utile conjointement au supervisé par exemple.

- Contraintes

Un jeu de données de ce qui est normal (connu).



Model1, Model2

Typical usecase

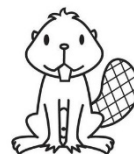


Model1 → It's a dog!

Model2 → It's normal!



Typical usecase 2



Model1 → It's a dog!

Model2 → It's an anomaly!



Incremental Learning (supervisé), continual learning, long-life learning



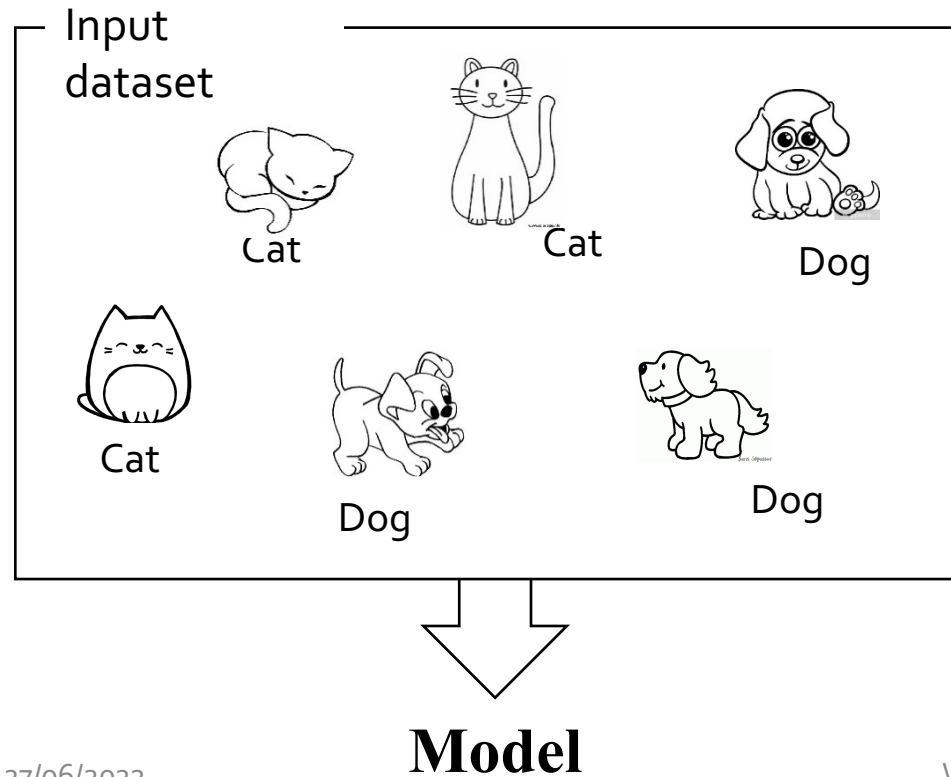
- Fonctions

Classification automatique.

- Contraintes

Un jeu de données labélisés, mais les données arrivent séquentiellement et l'ensemble des classes ne sont pas présente lors de l'apprentissage initial.

Step1:



Typical usecase



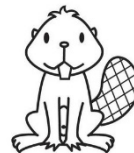
Model



It's a dog!



Typical usecase 2



Model



It's a dog!



Incremental Learning (supervisé), continual learning, long-life learning



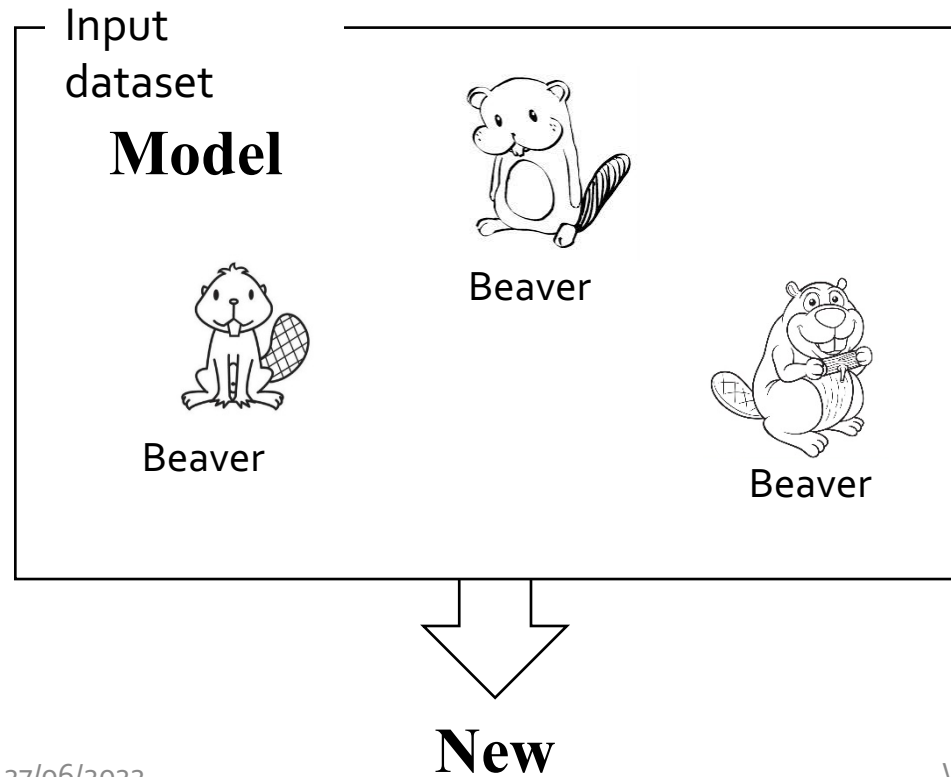
- Fonctions

Classification automatique. (ex : fonctionnement capteur autonome)

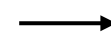
- Contraintes

Un jeu de données labélisés, mais les données arrivent séquentiellement et l'ensemble des classes ne sont pas présente lors de l'apprentissage initial.

Step2:



Typical usecase



New model



It's a dog!



Typical usecase 2



New model



It's a beaver!



Multimodalité & Fusion de données

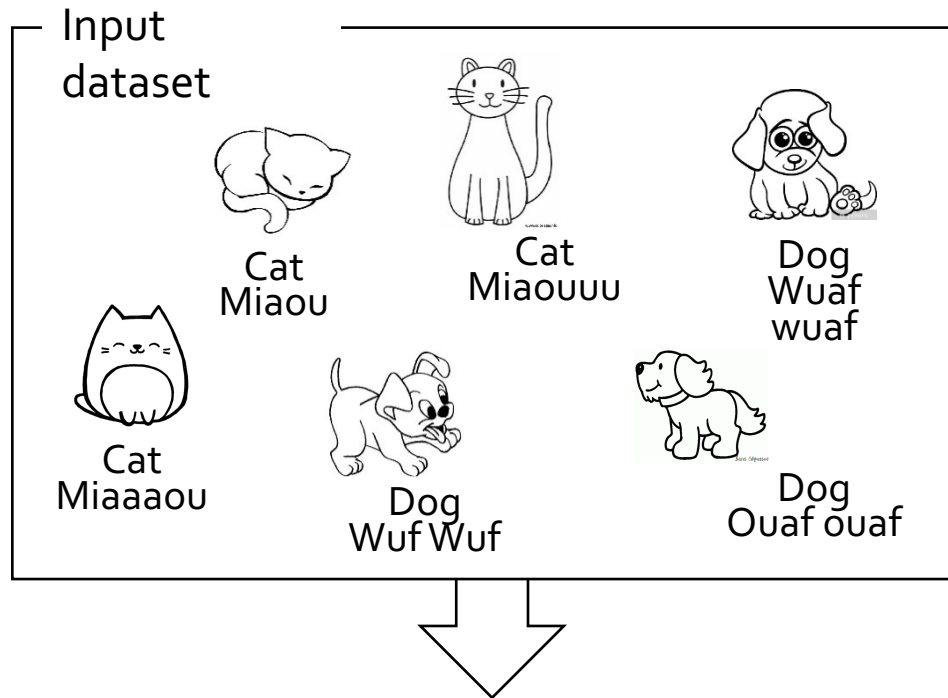
- Fonctions

Faire de l'apprentissage en tirant partie de plusieurs modalités, pour des obtenir des résultats + robustes.

(ex: image + son, modélisation feux de forets)

- Contraintes

Un jeu de données multimodal (labélisé pour du supervisé, non labélisé pour du non supervisé, etc).



More robust

Generative Adversarial Networks



NB : 1 GAN = 2 réseaux ... 2 fonctions

- Fonctions
 - Générer des données semblables aux données d'apprentissage.
 - Discriminer des données semblables, mais différentes des données d'apprentissages (fausses données).
- Contraintes
 - Un jeu de données.

Federated or collaborative learning



- Fonctions

Faire évoluer un modèle global à partir de plusieurs sources de données locales

- Contraintes

Un jeu de données provenant de plusieurs sources (labélisé pour du supervisé, non labélisé pour du non supervisé, etc).

- Principe

Peut être centralisé (utilisation d'un serveur central) ou décentralisé (les nœuds se synchronisent entre eux).



Image from <https://federated.withgoogle.com/>



Workshop Serenade - Malfante Malfante



M. Malfante, May 2020

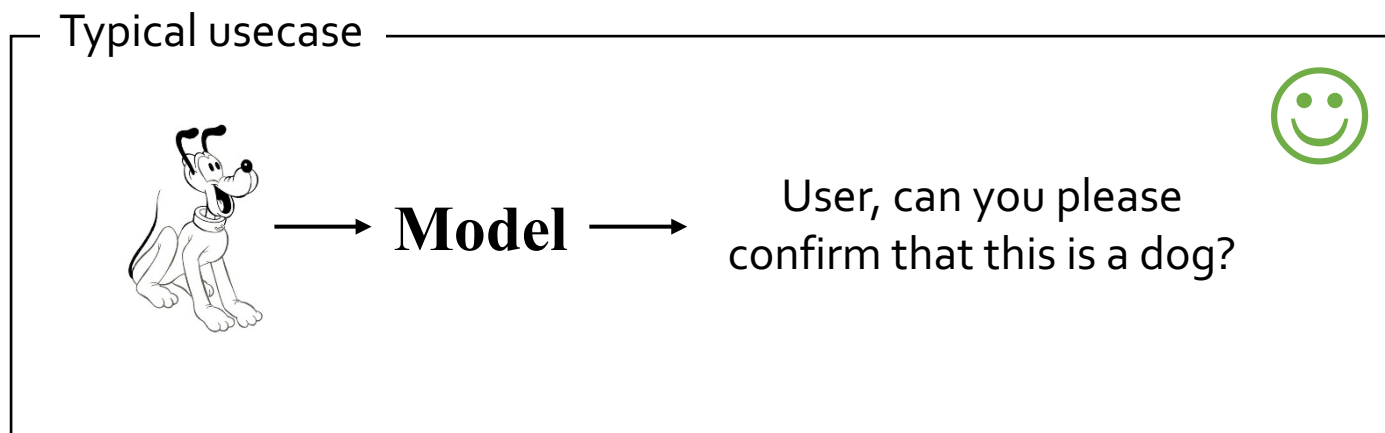
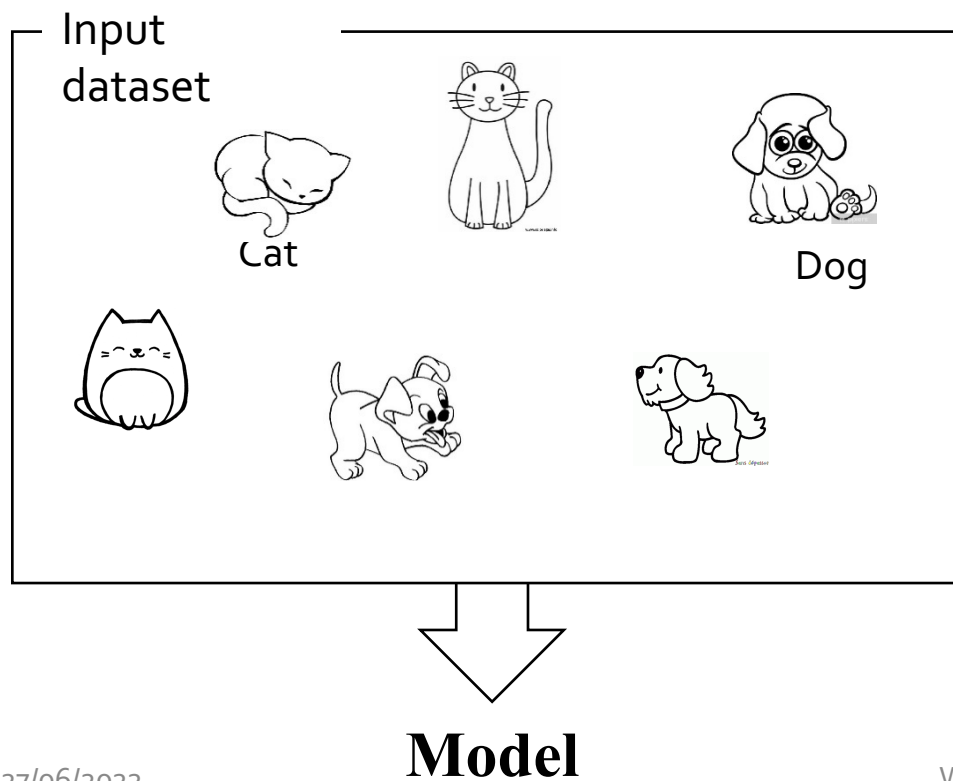
Active learning

- Fonctions

Les mêmes qu'en apprentissage supervisé (classification automatique)

- Contraintes

Quelques labels pour chaque classe (une grande partie des données n'est donc pas labélisée)



Partie 2 - Les Problèmes Que Je Souhaite Traiter Ne Sont Pas Dans Le Kit De Survie

- Théorie – Cartographie part II : les méthodes nouvelles
- Pratique & retour sur expérience : focus sur 5 points

#1 – Le cas particulier des séries temporelles.

Cas classique, classification automatique :

- Analyse de l'état de l'art
 - Etat de l'art images = CNN.
 - Etat de l'art séries temporelles = paroles = RNN
- Donc : je dois utiliser des RNN pour la classification automatique de séries temporelles.

Sauf que :

- Spécificité et multitudes de tâches associées au traitement de la parole (audio) et du langage (écrit)
.....
..... qui n'ont pas forcément grand-chose à voir avec les tâches de classification de séries temporelles.
- Spécificité des données de type texte et/ou parole.....
.....qui n'ont pas forcément grand-chose à voir avec les tâches de classification de séries temporelles.

>>> Les spécificités des données déterminent le choix de la méthode <<<

Donc : intéressant de s'inspirer des méthodes pour le traitement d'autres séries temporelles (comme la parole), comparer les spécificités des séries temporelles et identifier uniquement les briques pertinentes à réutiliser.

#2 - Focus on Representation Learning

- Transfer learning :
 - Train on dataset A
 - Use on dataset A'
- Self-supervised learning
 - Train on dataset A for task T
 - Use on dataset A for task T'
- Contrastive learning for representation learning
 - Train on dataset A for task T
 - Use on dataset A for task T'
- Contrastive learning for classification (end to end task)
 - Is new_data in the same class than this data ?
 - Nearest neighbour to find the closest data

#3 - IA & explicabilité



Observations <.....> Modèles .
=
Sciences

(Les modèles sont donc toujours faux)
(Modèles « physiques » : lois mécaniques par exemple)

Quid de l'IA là dedans ?

- Intérêt quand on n'arrive pas à trouver la loi qui régit les données (à construire de modèle physique)
- L'IA peut s'inscrire dans une démarche de compréhension : IA explicables, IA pour expliquer

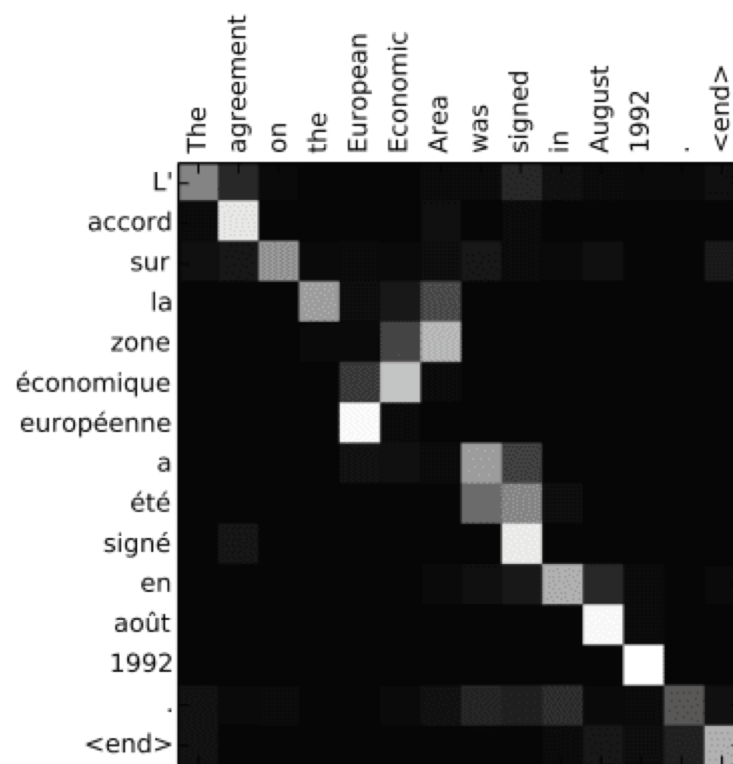
#3 a) IA explicables

Explicables < > Fiables

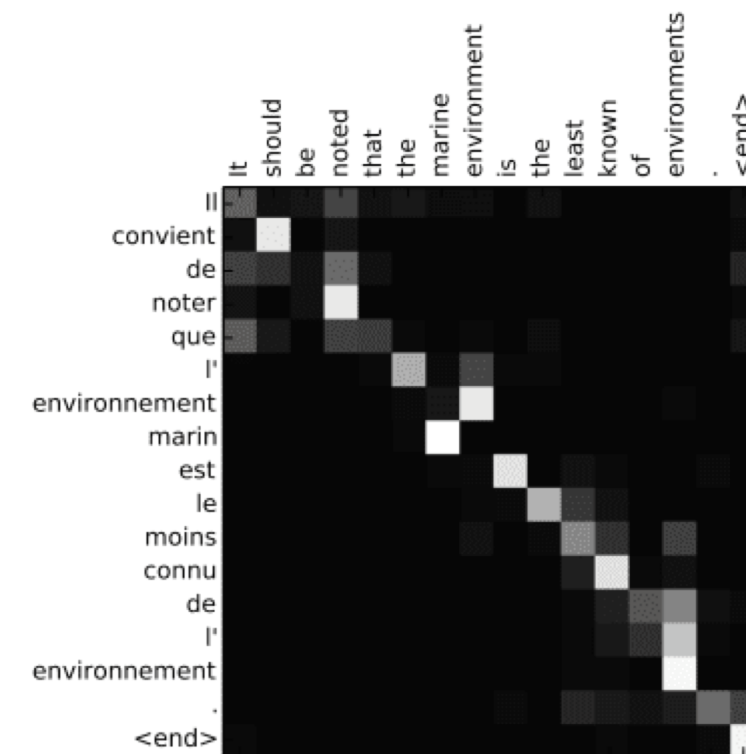
Pipeline supervisé
Recordings

Méthodes explicables par des

- Disentangled representations
- Case based reasoning
- Concept based
- Attention based models
- Régularisation



(a)



(b)

Attentional Interpretation of French to English Translation, Bahdanau, et al.

#3 b) IA pour l'explicabilité



Soit :

- Utiliser l'IA pour faire une analyse qui permet de comprendre le phénomène
- Utiliser l'IA pour modéliser un phénomène, puis comprendre le modèle appris pour comprendre le phénomène étudié back to explainable AI.

Exemples d'études / d'outils :

- Étude des représentations apprises
 - Visualisation de features
 - « Discovering physical concepts with neural networks »
 - Etude de la position d'un pendule oscillant.
 - Input: positions passées
 - Output: prédiction à l'instant suivant.
- Mécanismes à attention ?
 - « Attention is not explanation »
 - « Attention is not not explanation »

#4 - L'IA et la fiabilité



Fiabilité :

- Lien avec l'explicabilité, mais pas que
- Les réseaux de neurones surestiment les probabilités de prédiction, modélisation d'incertitude (soit dans le modèle, soit via la labélisation)
- Détection d'anomalies. Problème de la validation.

#5 – L'ordre des données dans les batchs

Apprentissage par batch Comment construire les batchs ?

Usecase typiques :

- Apprentissage incrémental
- Jeux de données déséquilibrés
- Curriculum learning

Take home messages, Discussions

- La communauté IA développe des outils, sur des usecases standards.
- Quels usages ? Quels impacts ?
 - L'IA, oui mais là où il faut.
 - L'IA pour l'applicatif. Le besoin est déjà là.
 - Pas nécessairement besoin de créer un nouveau besoin.

CONTACT INFO

@CEA.FR

(PROJETS EN COURS
DE MONTAGE SUR L'IA
EMBARQUÉ, BESOIN DE
USECASES :
[HTTPS://ANR.FR/AMI-](https://anr.fr/AMI-)