



DÉVELOPPER LA PENSÉE CRITIQUE, GRÂCE À L'IA

...POUR MIEUX CRITIQUER L'IA ?



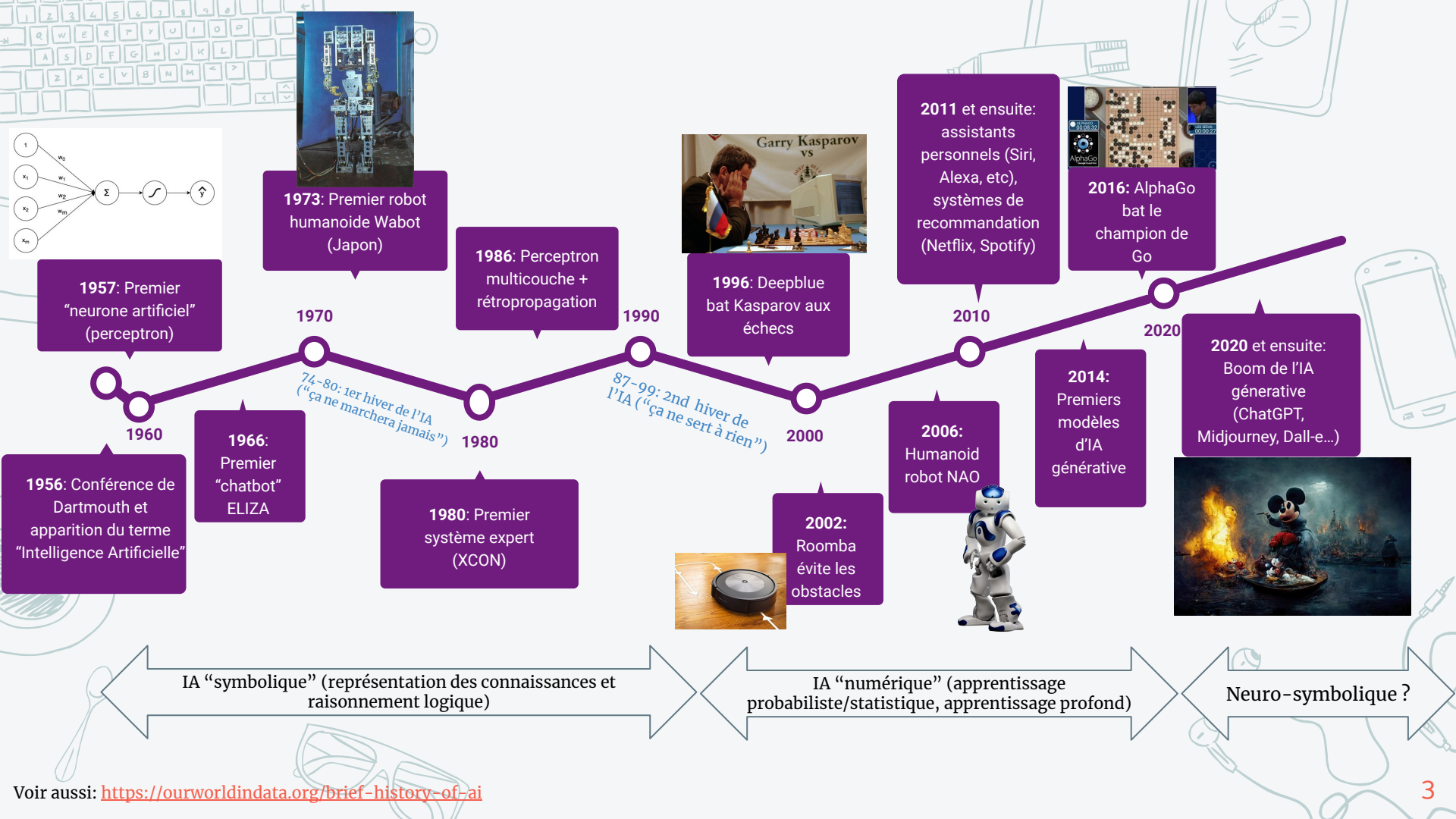
Marvin Minsky, 1968

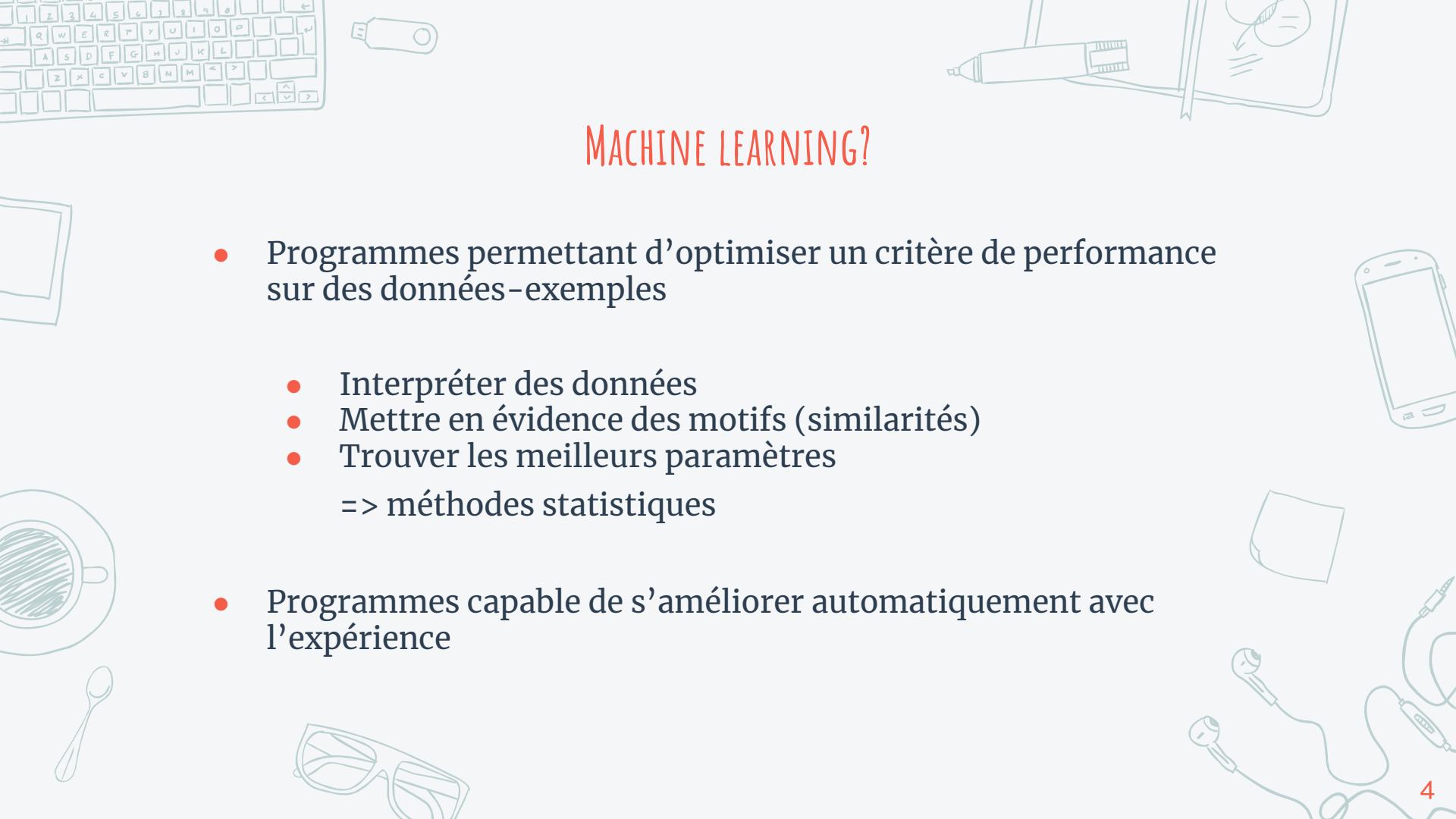
IA: *“the science of making machines do things that would require intelligence if done by human”*

Machine Learning: *“the field of study that gives computers the ability to learn without being explicitly programmed”*



Arthur Samuel, IBM, 1959



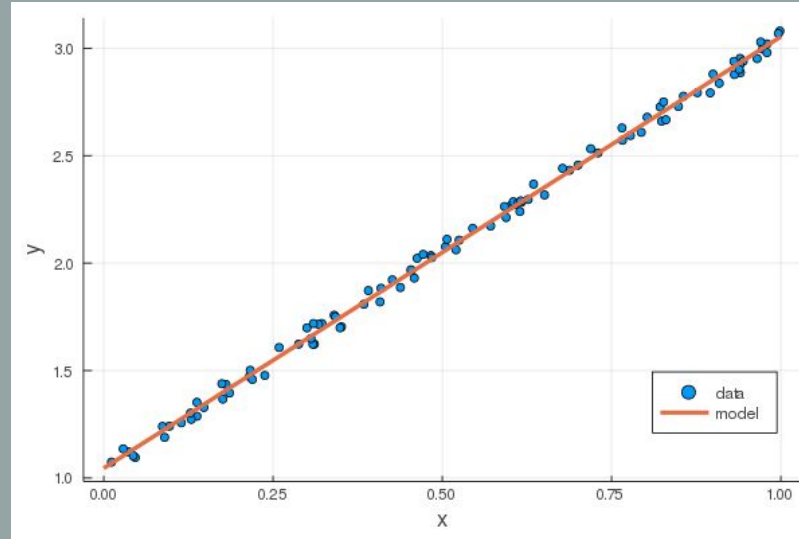


MACHINE LEARNING?

- Programmes permettant d'optimiser un critère de performance sur des données-exemples
 - Interpréter des données
 - Mettre en évidence des motifs (similarités)
 - Trouver les meilleurs paramètres=> méthodes statistiques
- Programmes capable de s'améliorer automatiquement avec l'expérience

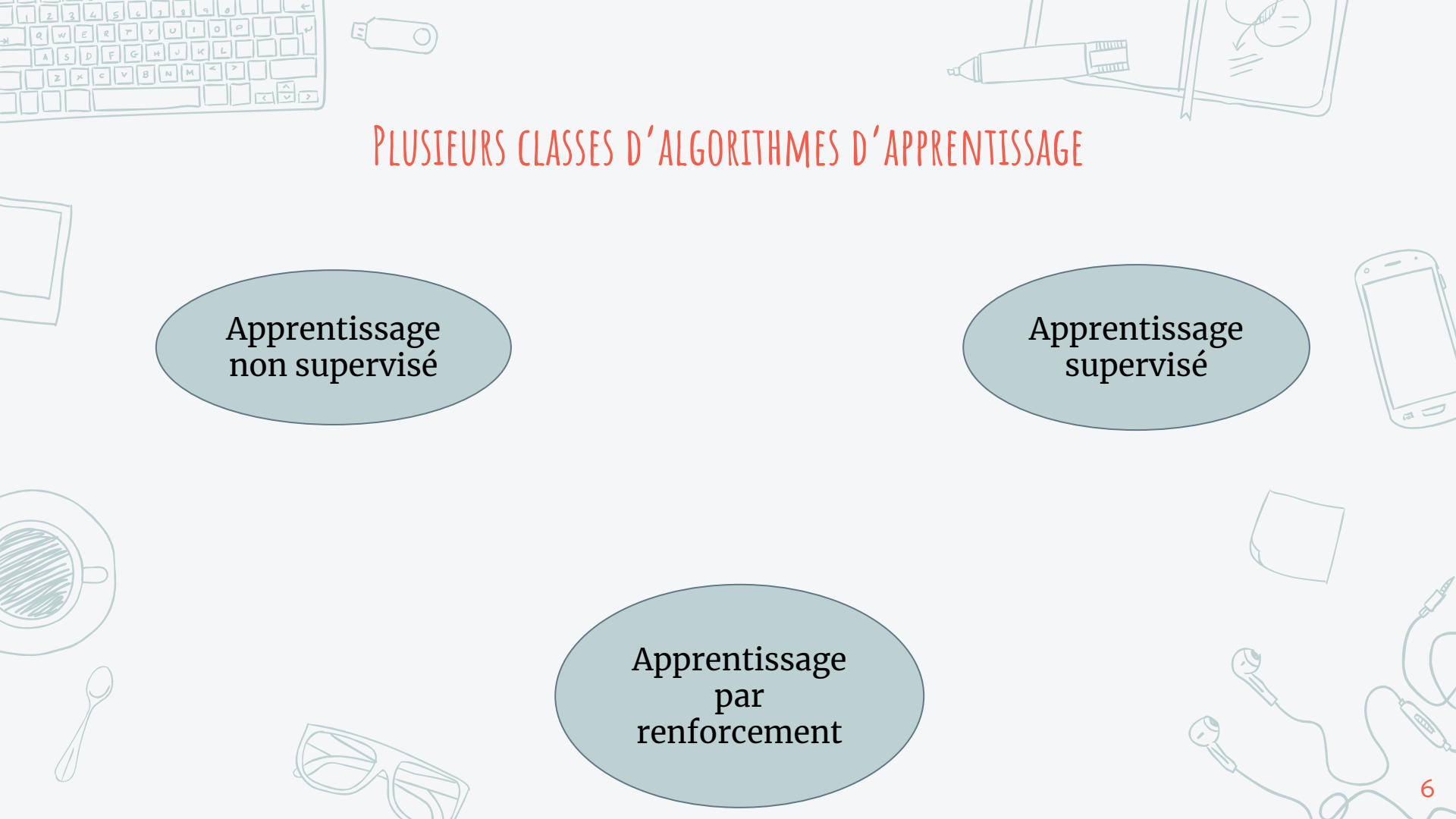
Théorie de la probabilité
⇒ prédire de nouvelles
observations

Théorie:
Modèles



Evidence
empirique:
Données

Statistiques
⇒ décrire/expiquer les
données observées,
⇒ i.e. inférer et ajuster
les paramètres du
modèle



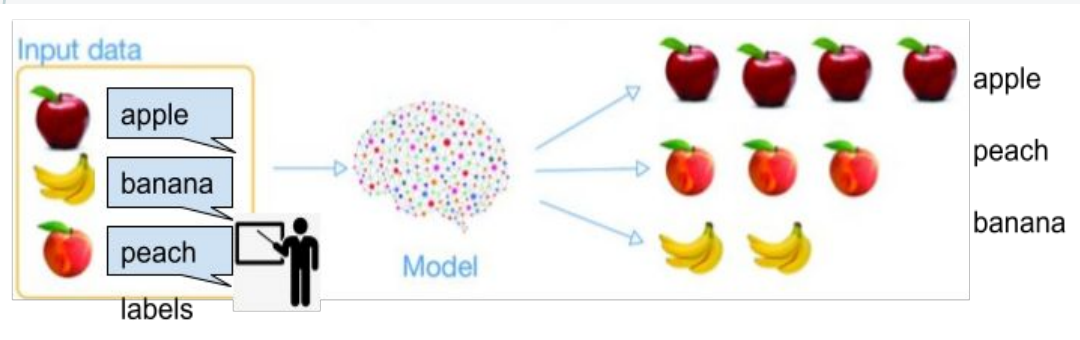
PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE

Apprentissage
non supervisé

Apprentissage
supervisé

Apprentissage
par
renforcement

PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE



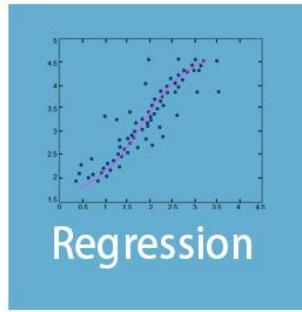
Apprentissage
supervisé

Données-exemples étiquetées

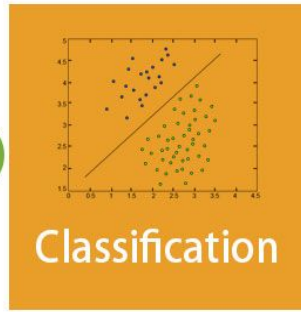
⇒ Objectif: minimiser l'erreur.



PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE



vs



What will be the temperature tomorrow?

84°

Fahrenheit



Will it be hot or cold tomorrow?

COLD

HOT

Fahrenheit

Apprentissage supervisé

⇒ Classification

Ex: Classifieur Bayésien naïf, Arbres de décision, certains réseaux de neurones artificiels (eg, réseaux à convolution pour la reconnaissance d'images)

⇒ Régression

Ex: Régression linéaire, logistique

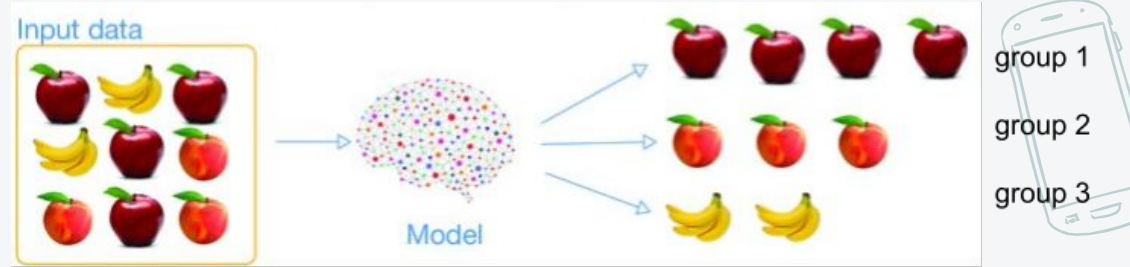
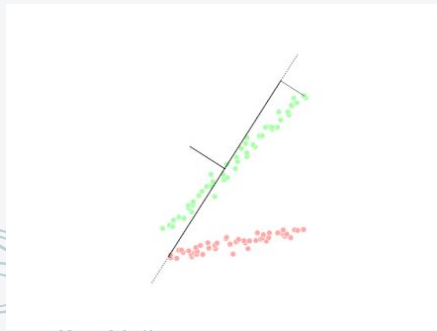
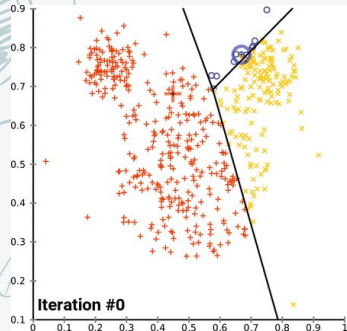


PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE

Apprentissage
non supervisé

Données-exemples non étiquetées

Regroupement (*clustering*) ou Réduction de la dimension



⇒ Objectif: regrouper ensemble les éléments “proches”
(calcul d’une distance).



PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE

Apprentissage non supervisé

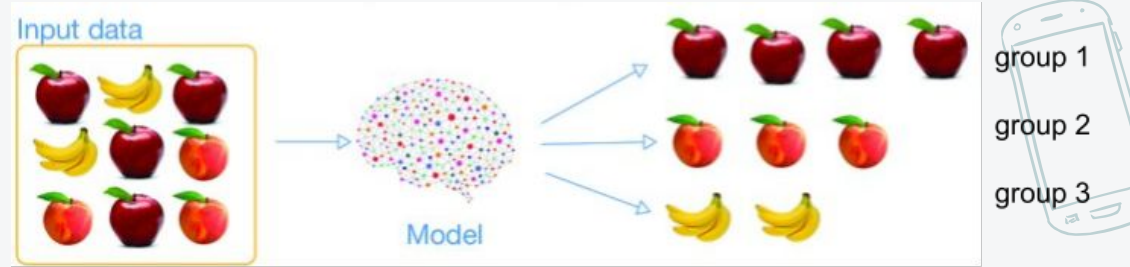
Données-exemples non étiquetées

⇒ Regroupement (*clustering*)

Ex: K-moyennes, modèles de mélange...

⇒ Réduction de la dimensionnalité

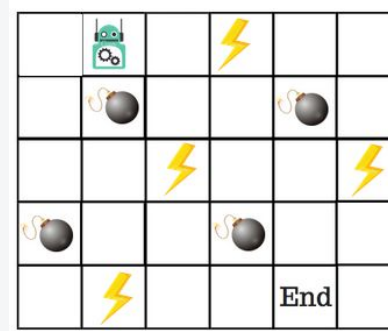
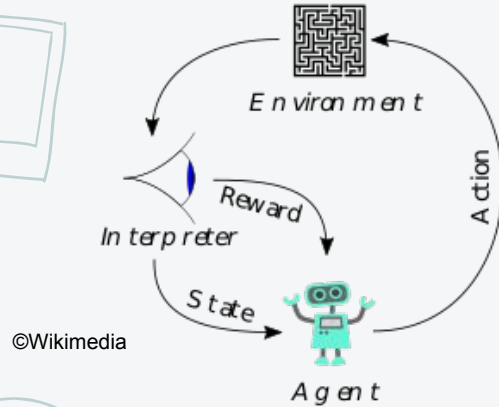
Ex: Analyse par composantes principales



⇒ Objectif: regrouper ensemble les éléments "proches" (calcul d'une distance).



PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE



Actions : ↑ → ↓ ←

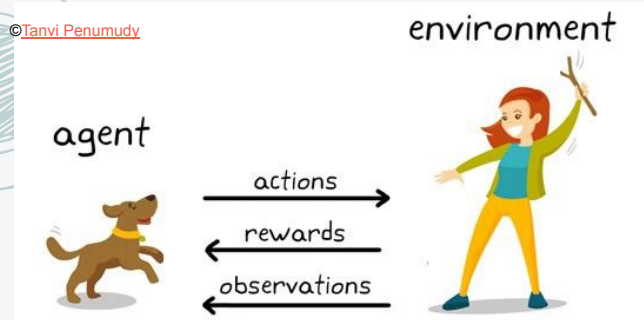
Start	0	0	0	0
Nothing / Blank	0	0	0	0
Power	0	0	0	0
Mines	0	0	0	0
END	0	0	0	0

©ADL.freecodecamp

Ex: une table de Q-valeurs.

- **recharge** = +1
- **mine** = -100
- **fin** = +100

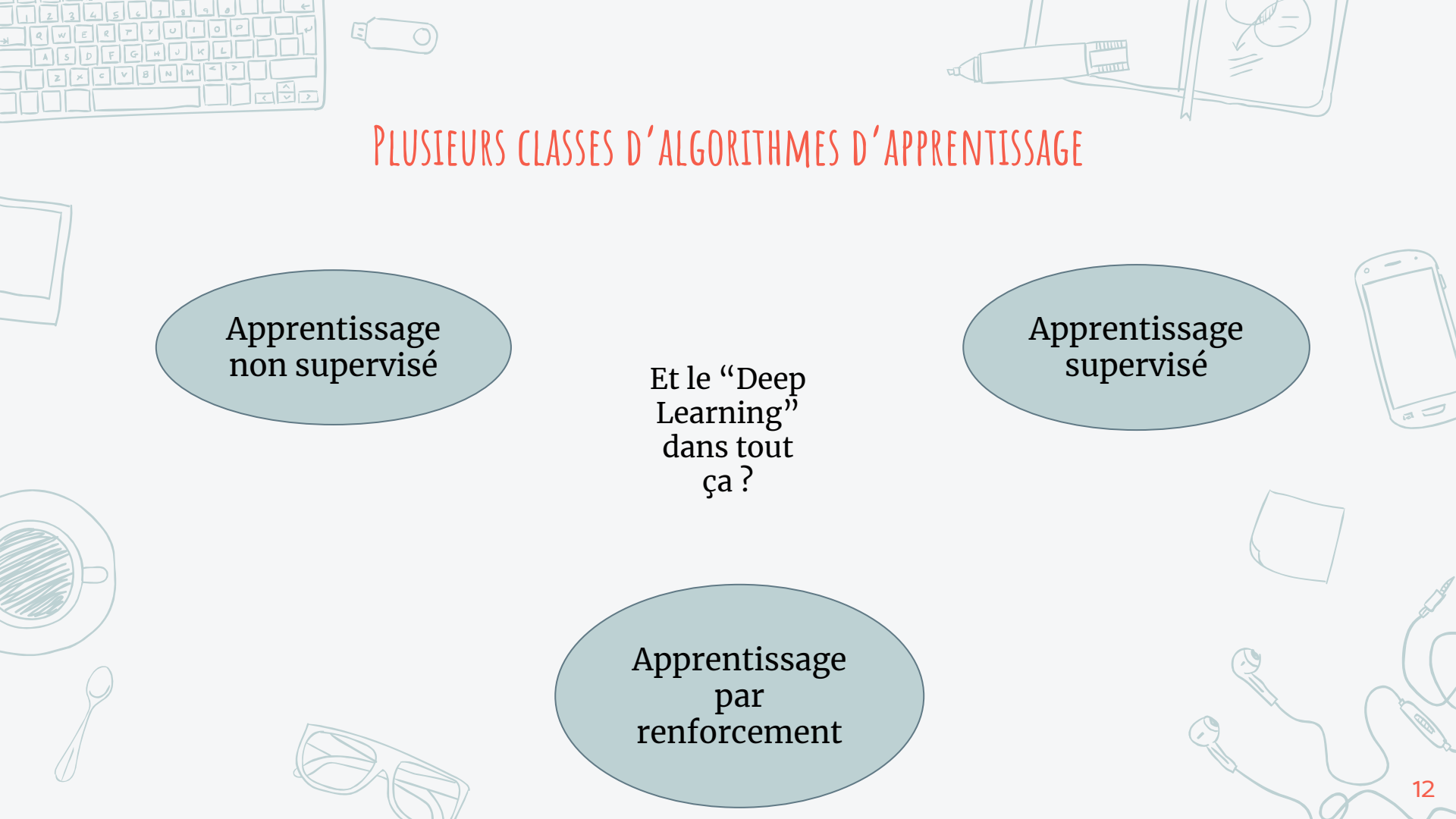
A chaque étape, on calcule la Q-valeur d'un couple (état, action) en fonction de la récompense obtenue.



Apprentissage
par
renforcement

⇒ Objectif: Choisir les actions qui maximisent la récompense à plus ou moins long terme.

Ex: AlphaGo (deep RL)



PLUSIEURS CLASSES D'ALGORITHMES D'APPRENTISSAGE

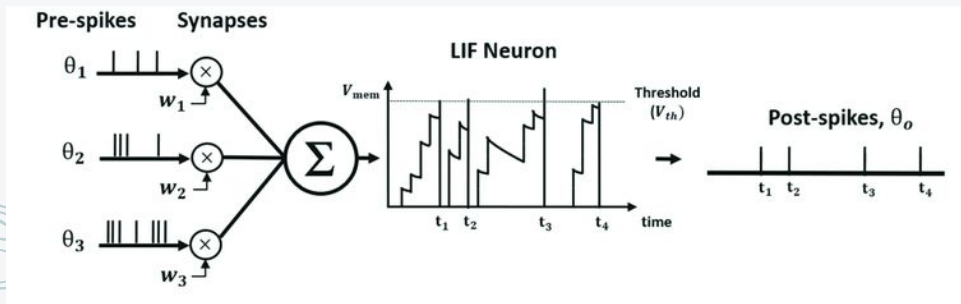
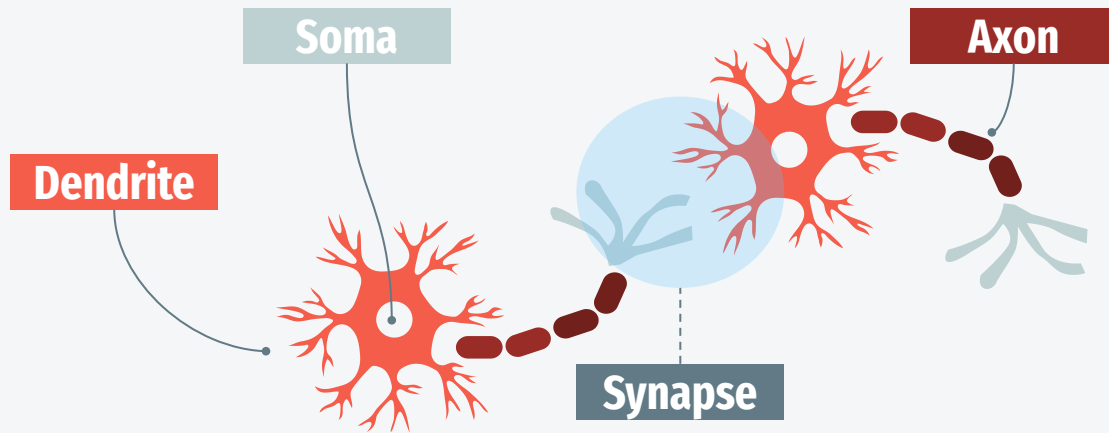
Apprentissage
non supervisé

Et le “Deep
Learning”
dans tout
ça ?

Apprentissage
supervisé

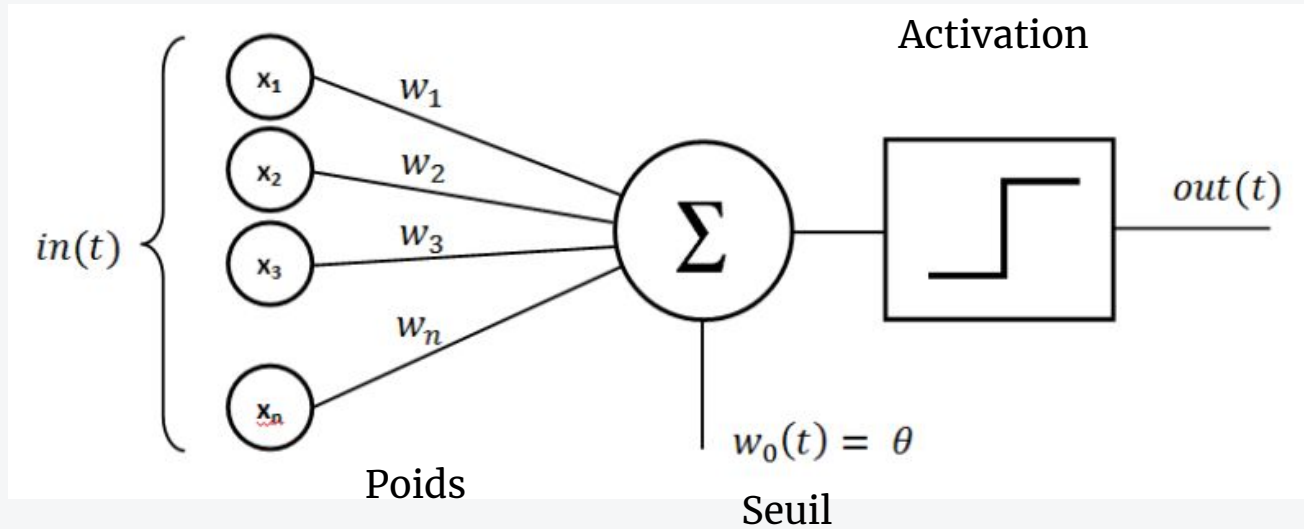
Apprentissage
par
renforcement

NEURONE BIOLOGIQUE



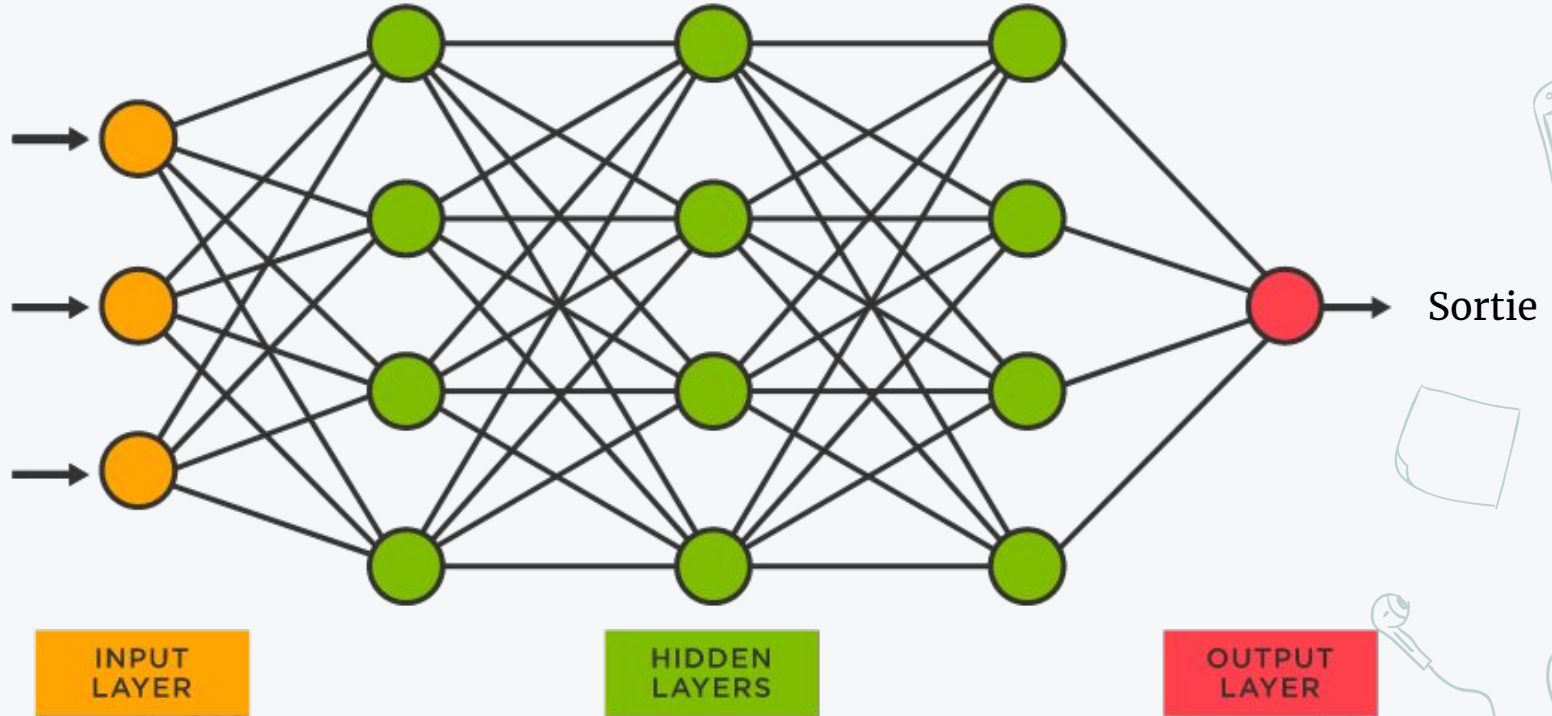
NEURONES ARTIFICIELS

Entrée



Un neurone artificiel (perceptron)

RÉSEAU DE NEURONES



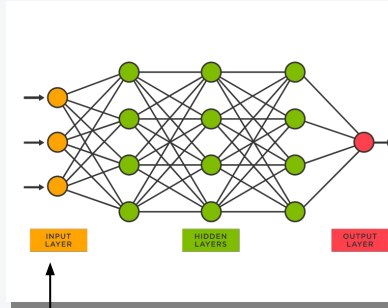
ENTRAÎNER LE RÉSEAU (SUPERVISÉ)

Entrée

0	4	1	9	2	1	3	1	4	3
5	3	6	1	7	2	8	6	9	4
0	9	1	1	2	4	3	2	7	3
8	6	9	0	5	6	0	7	6	1
8	7	9	3	9	8	5	9	3	3
0	7	4	9	8	0	9	4	7	4
4	6	0	4	5	6	1	0	0	1
7	1	6	3	0	2	1	1	7	9
0	2	6	7	8	3	9	0	4	6
7	4	6	8	0	7	8	3	1	5

MNIST
dataset

ANN (modèle)



Sortie du
modèle

0 1 1 9 2 1 3 1 4 3
5 3 6 1 7 2 8 6 9 4
0 9 1 1 2 4 3 2 7 3
...

Mettre à jour les
poids

(par exemple ici: rétropropagation
du gradient*)

Calcul de l'erreur

Sortie
attendue

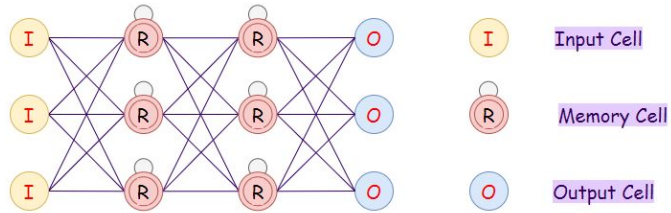
0 4 1 9 2 1 3 1 4 3
5 3 6 1 7 2 8 6 9 4
0 9 1 1 2 4 3 2 1 3
...

Minimiser
l'erreur

*Il existe d'autres types d'entraînement sans rétropropagation, comme dans les réseaux de neurones à réservoirs.

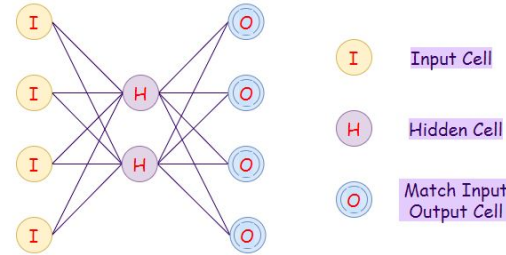
DIFFÉRENTES ARCHITECTURES POUR DIFFÉRENTS PROBLÈMES

Long/Short Term Memory (LSTM)



LSTM pour les séries temporelles

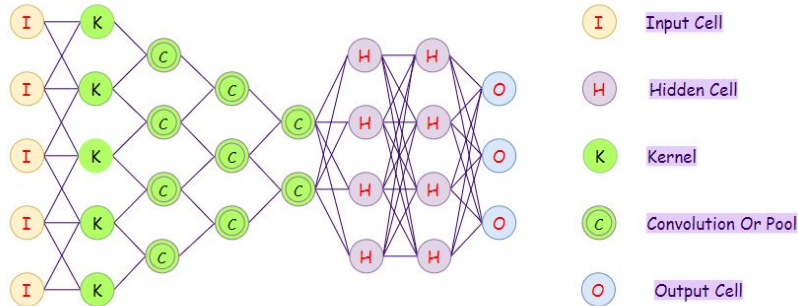
Auto Encoder (AE)



AutoEncodeur pour la classification

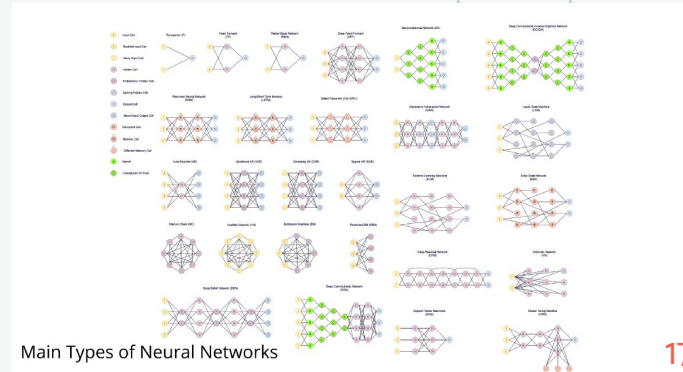
CNN pour la vision

Deep Convolutional Network (DCN)



etc!

cf: [Main Types of Neural Networks and Applications](#)



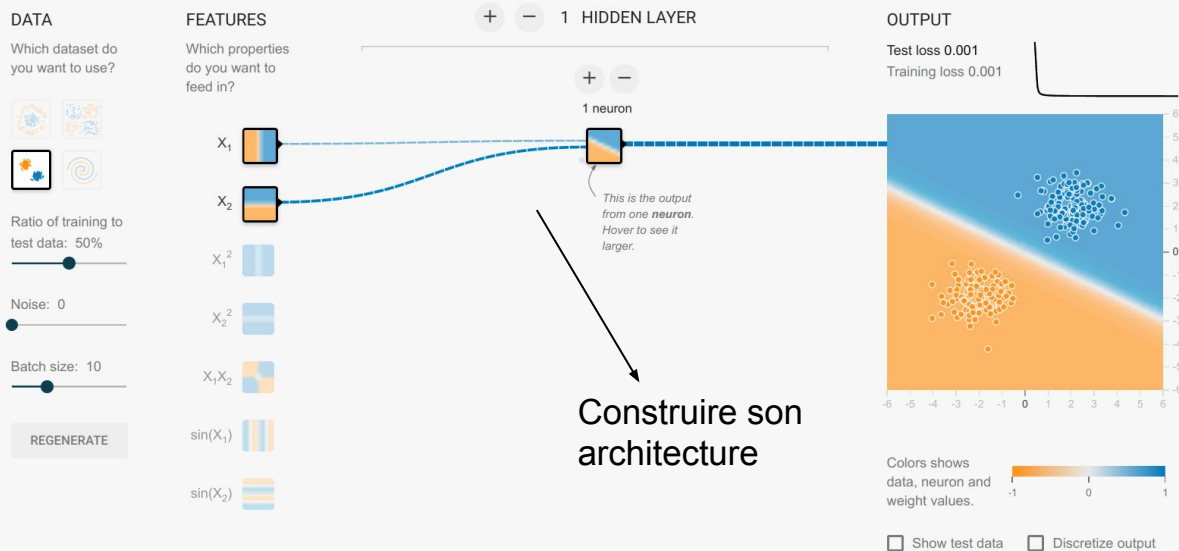
Main Types of Neural Networks

COMMENT CHOISIR UNE ARCHITECTURE?

Lancer
l'entraînement

🔄 ▶ 000,158 Learning rate: 0.03 Activation: Tanh Regularization: None Regularization rate: 0 Problem type: Classification

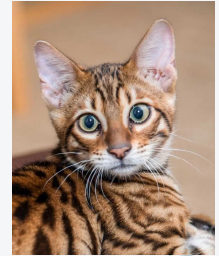
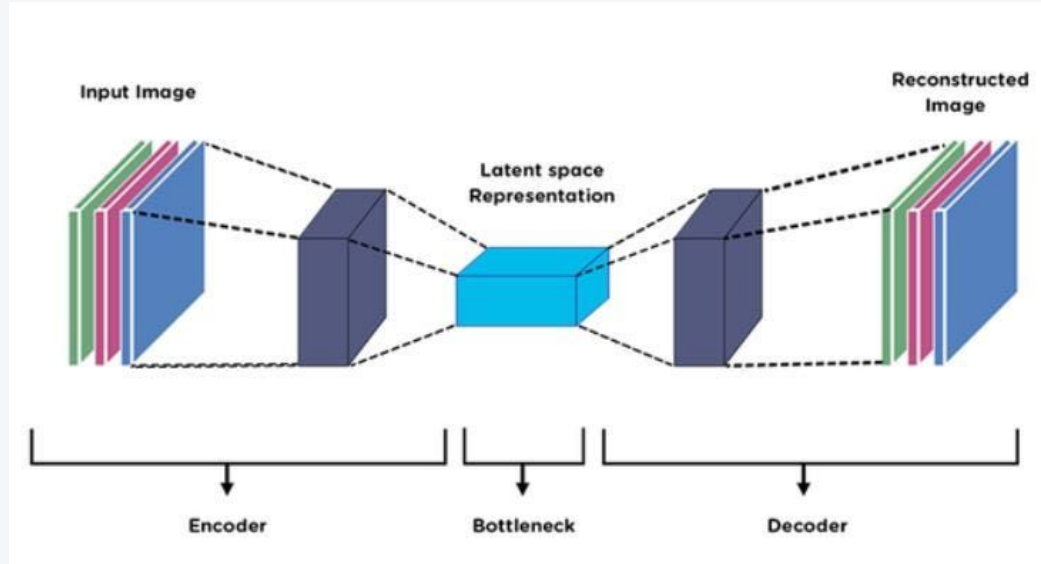
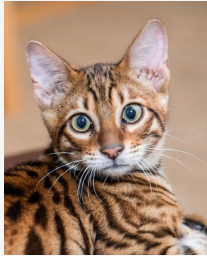
Choisir les
hyper-paramètres
(paramètres qui ne
seront pas mis à jour
pendant l'entraînement)



Construire son
architecture

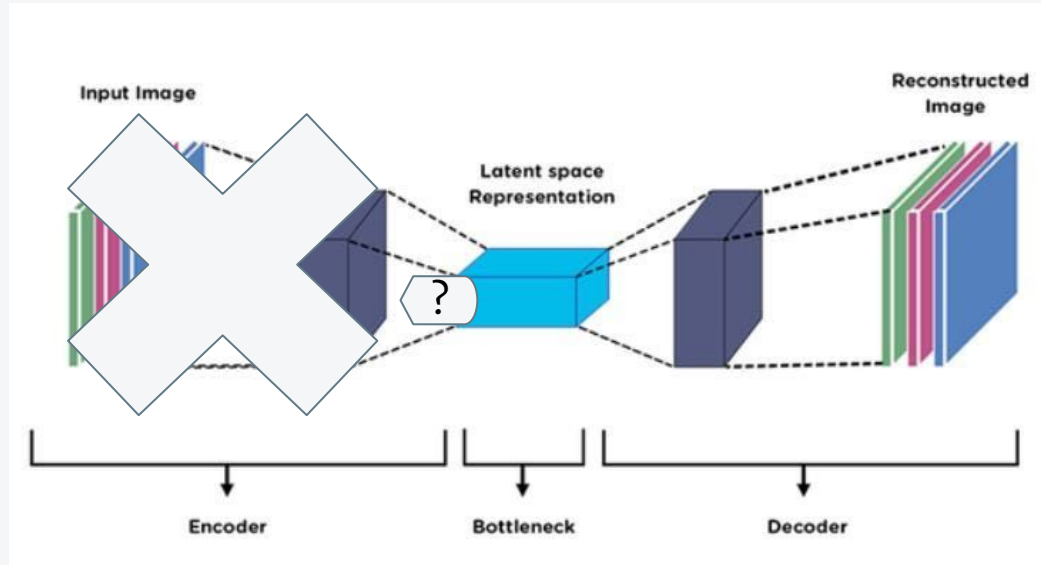
A tester sur playground.tensorflow.org :)
(exemple d'activité guidée [ici](#))

L'IA GÉNÉRATIVE: PRINCIPE



Encodeur + décodeur: à la base du “transformer”

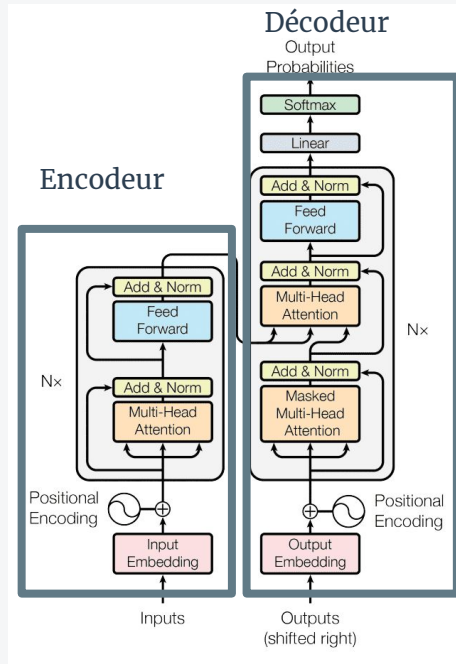
L'IA GÉNÉRATIVE: PRINCIPE



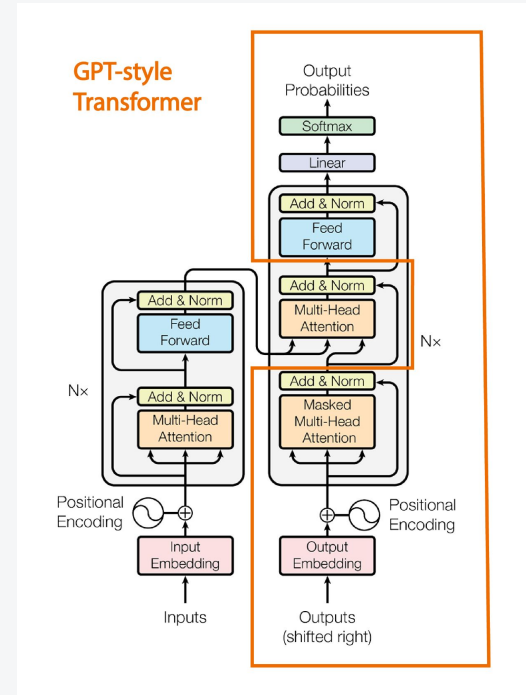
Génération possible par le décodeur après pré-entraînement



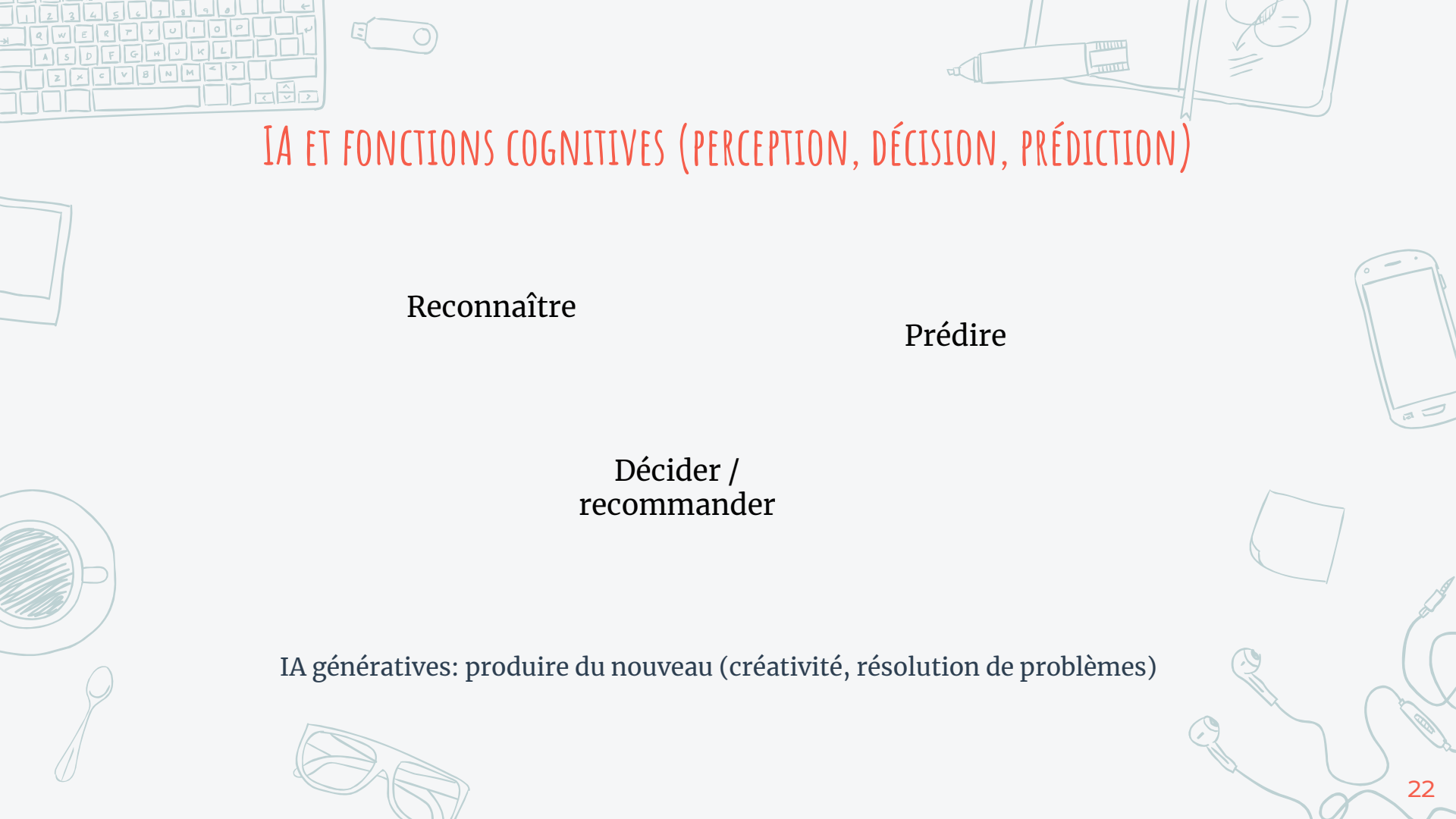
L'IA GÉNÉRATIVE: DU TRANSFORMER AU GPT



Encodeur + décodeur: à la base du “transformer”
Cf: [\[1706.03762\] Attention Is All You Need](#)



Le GPT “generative pretrained transformer”: on entraîne l’ensemble mais on garde la partie décodeur pour la génération
Cf: [The Transformer architecture of GPT models](#)



IA ET FONCTIONS COGNITIVES (PERCEPTION, DÉCISION, PRÉDICTION)

Reconnaître

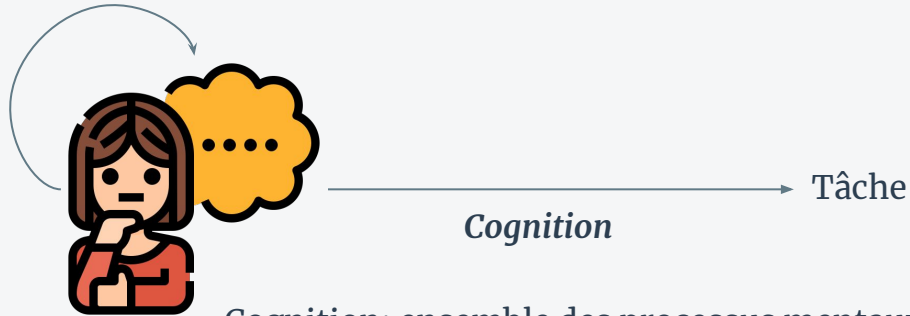
Prédire

Décider /
recommander

IA génératives: produire du nouveau (créativité, résolution de problèmes)

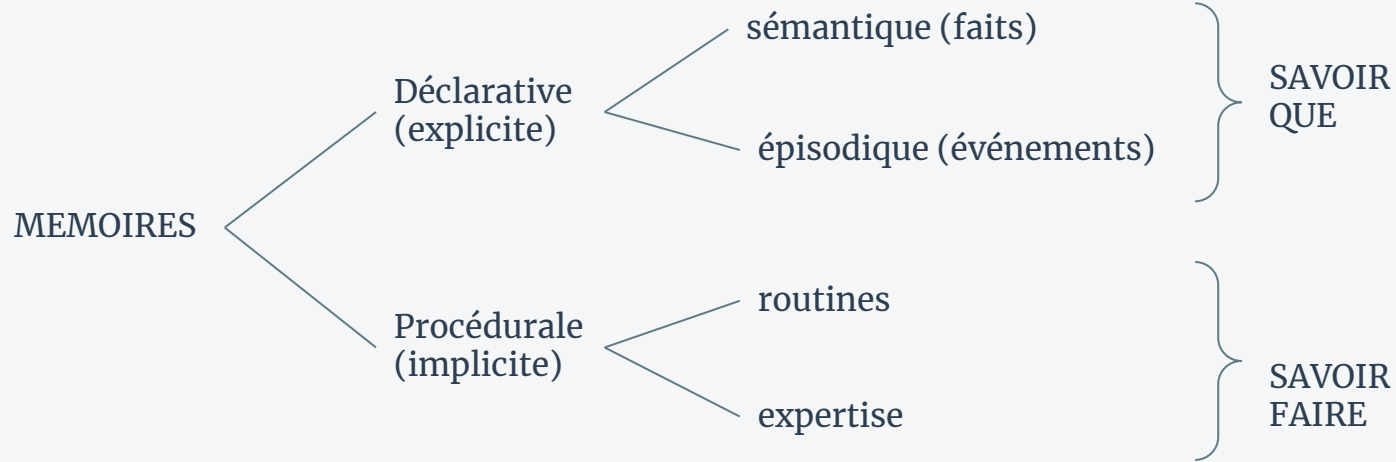
COGNITION ET METACOGNITION

Métacognition:
Réflexion sur ses propres
processus mentaux

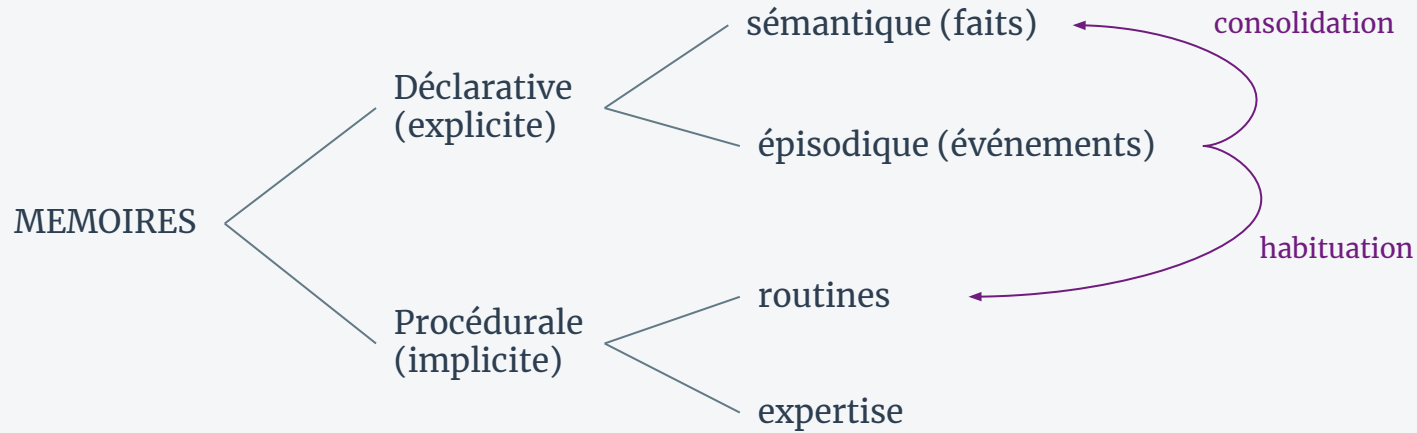


Cognition: ensemble des processus mentaux, explicites et implicites, tel que: la mémoire, la perception, l'attention, la motricité, les émotions, le langage, le raisonnement, l'apprentissage, la prise de décision, la planification... y compris les fonctions de haut niveau (résolution de problème, créativité) et les processus métacognitifs.

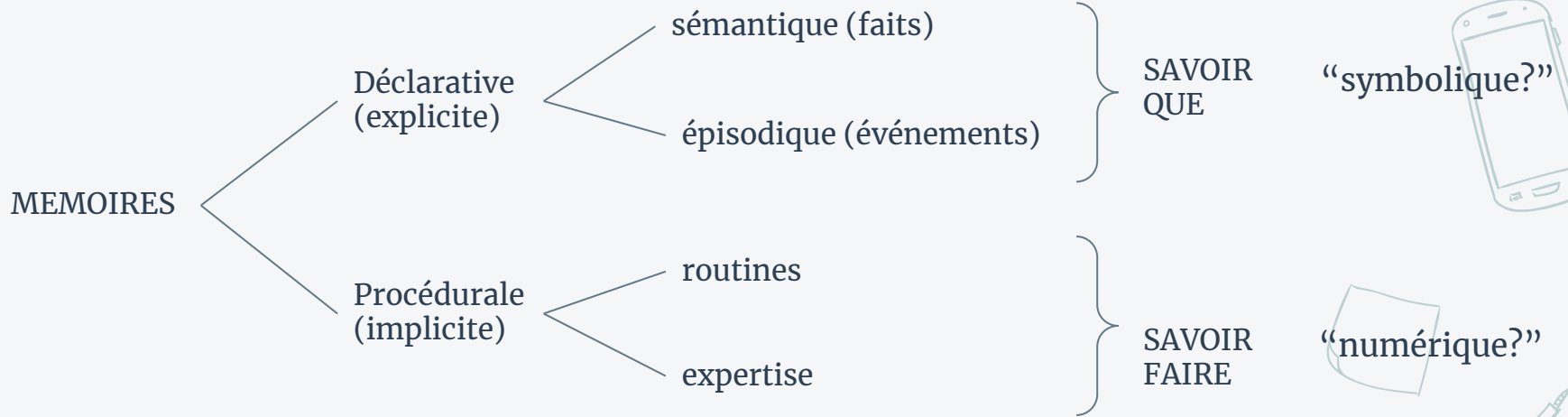
DIFFÉRENTS TYPES DE MÉMOIRES



DIFFÉRENTS TYPES DE MÉMOIRES EN INTERACTION



DIFFÉRENTS TYPES DE MÉMOIRES... EN LIEN AVEC DIFFÉRENTS TYPES D'IA?



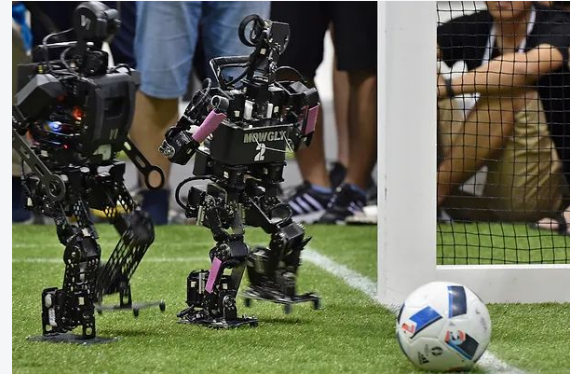
Voir aussi: [Frédéric Alexandre - Le cerveau est-il un bon modèle d'Intelligence Artificielle ?](#)

QUELQUES MOTS SUR LA COGNITION INCARNÉE

Peut-on réellement comprendre les concepts sans aucune expérience sensible du monde ?

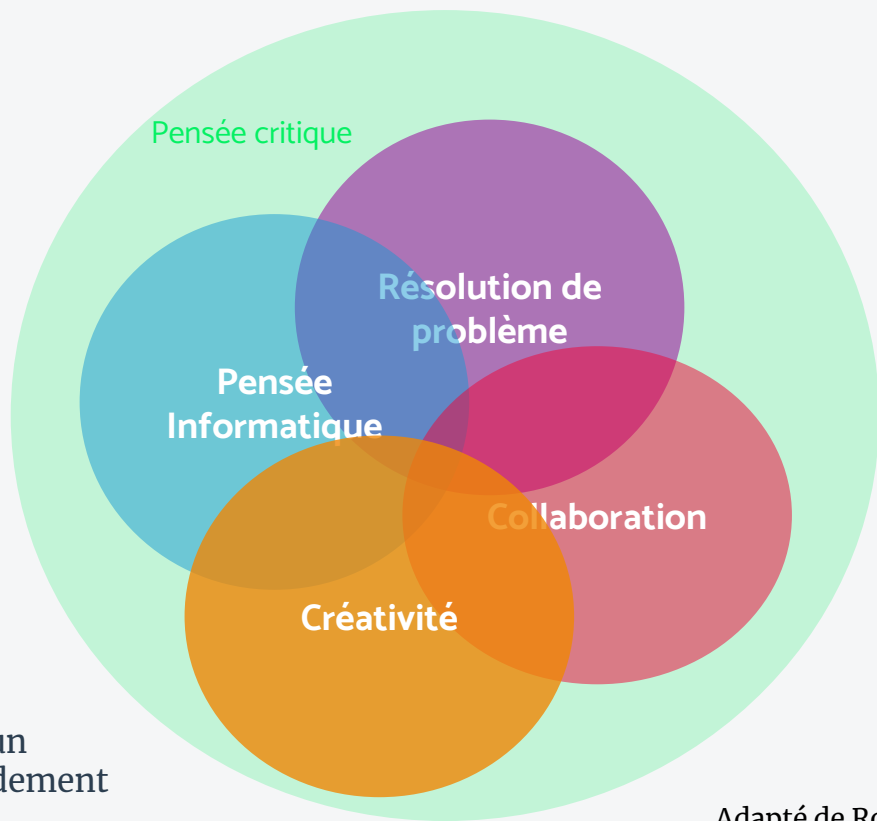
Voir aussi [L'expérience sensible du monde](#) | NICOLAS ROUGIER

Et la conversation entre Stevan Harnad et ChatGPT à ce sujet: [\[24.02.02243\] Language Writ Large: LLMs, ChatGPT, Grounding, Meaning and Understanding](#)



Rhoban le groupe de recherche en robotique autonome du LaBRI

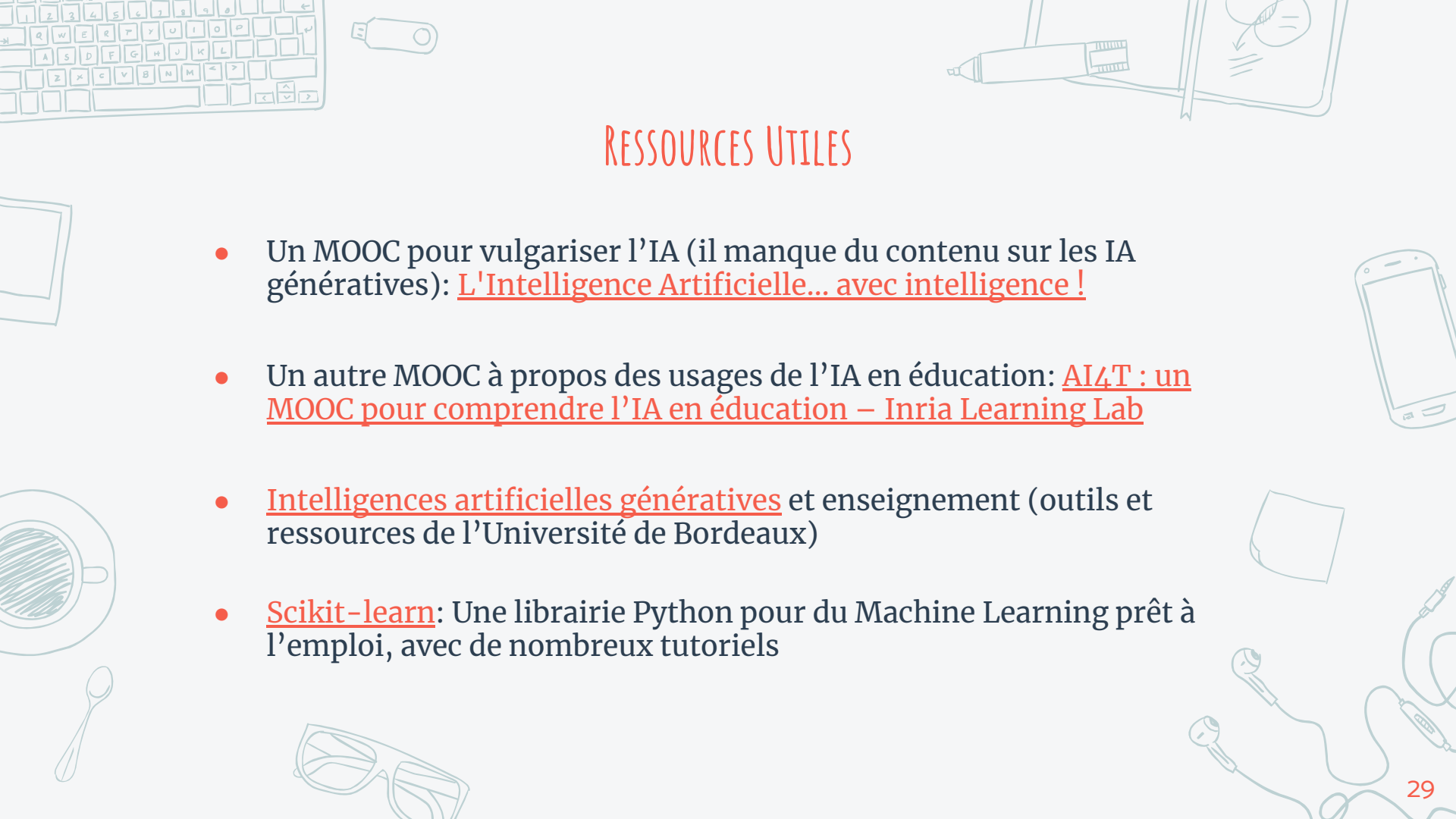
COMPÉTENCES TRANSVERSALES: KEZAKO?



STEAM
= science, technology,
engineering, art and
mathematics

Nécessité de s'adapter dans un
monde évoluant (trop?) rapidement

Adapté de Romero (2016),
cf: [#5c21 : 5 compétences clés pour le 21e siècle](#)



RESSOURCES UTILES

- Un MOOC pour vulgariser l'IA (il manque du contenu sur les IA génératives): [L'Intelligence Artificielle... avec intelligence !](#)
- Un autre MOOC à propos des usages de l'IA en éducation: [AI4T : un MOOC pour comprendre l'IA en éducation – Inria Learning Lab](#)
- [Intelligences artificielles génératives](#) et enseignement (outils et ressources de l'Université de Bordeaux)
- [Scikit-learn](#): Une librairie Python pour du Machine Learning prêt à l'emploi, avec de nombreux tutoriels