



**HAL**  
open science

# Exploration des Transformers

Yannis Bendi-Ouis, Xavier Hinaut

► **To cite this version:**

Yannis Bendi-Ouis, Xavier Hinaut. Exploration des Transformers. Master. France. 2023. hal-04850985

**HAL Id: hal-04850985**

**<https://hal.science/hal-04850985v1>**

Submitted on 20 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

*Inria*

Exploration des  
Transformers



LLAMA 2



CHATGPT



MISTRAL AI

*Inria*

# Yannis Bendi-Ouis

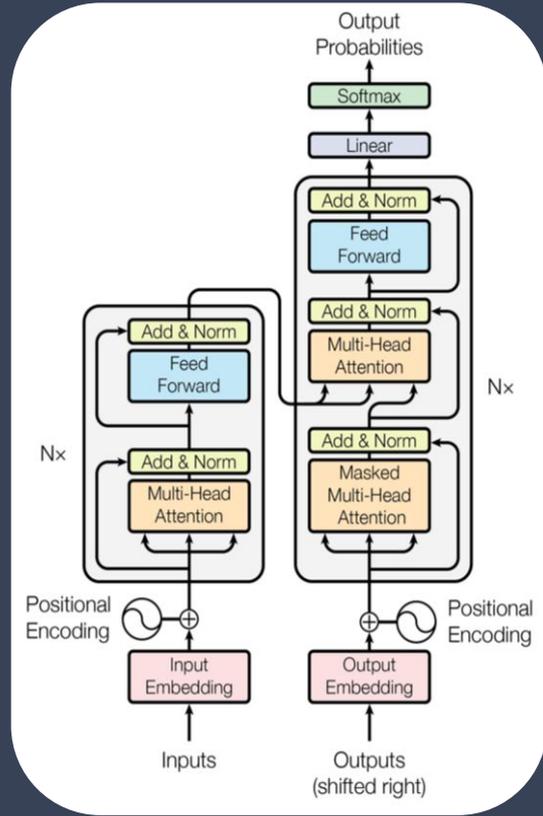
Doctorant en Intelligence Artificielle / Neuroscience

[yannis.bendi-ouis@inria.fr](mailto:yannis.bendi-ouis@inria.fr)

[www.naowak.fr](http://www.naowak.fr)



# Architecture d'un Transformer



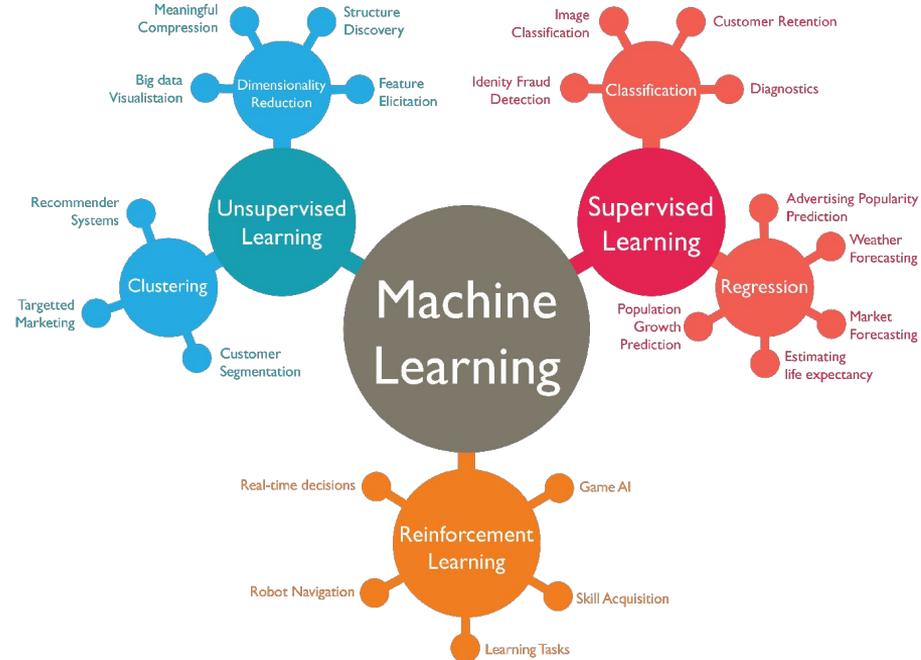
## Sommaire

01. Qu'est ce que le Machine Learning ?
02. Comment fonctionne un réseau de neurones artificiels ?
03. Comment faire avec des séquences ?
04. Qu'est ce qu'un Transformers ?
05. Comment entraîner un Transformers ?
06. Qu'est ce qu'un prompt ?
07. Quels sont les risques ?
08. Débat / Questions

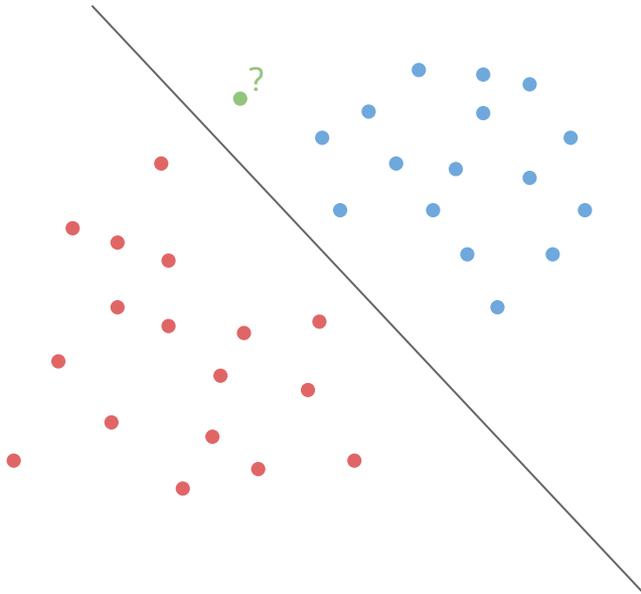
01

Qu'est ce que le  
machine learning ?

# Une grande famille d'algorithmes

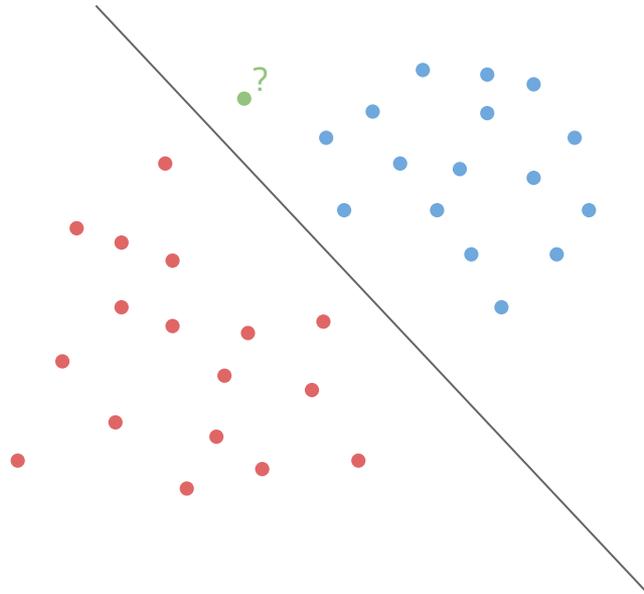


# Classification / Régression



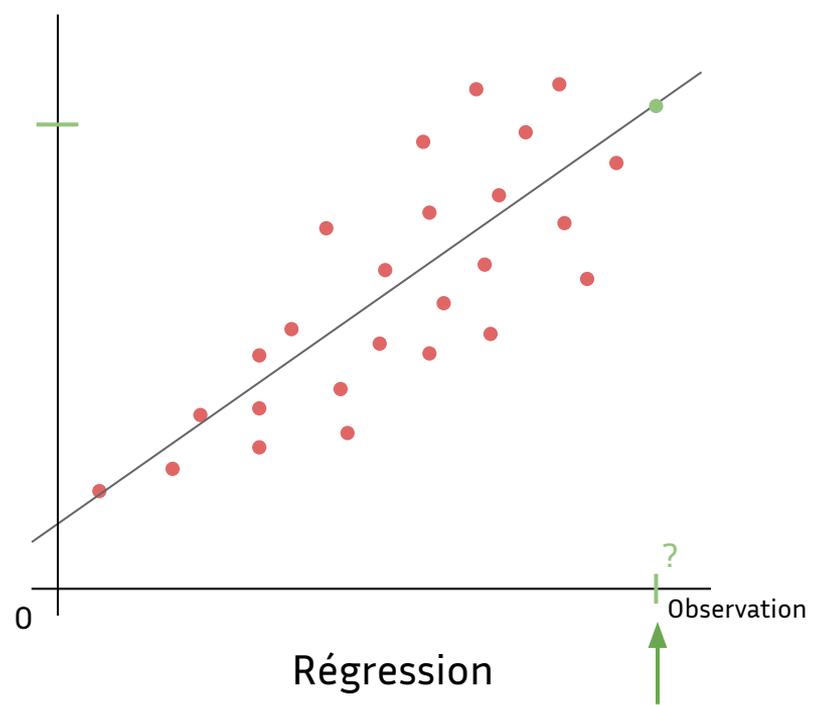
Classification

# Classification / Régression



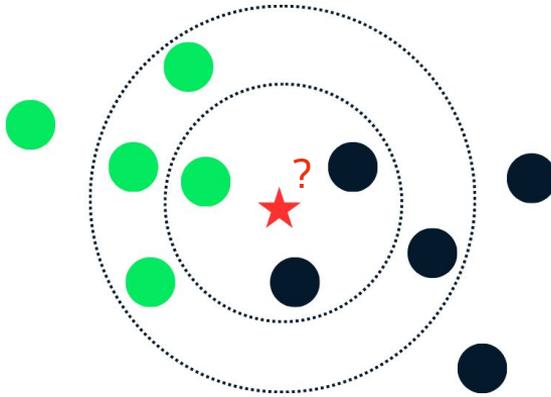
Classification

Prédiction



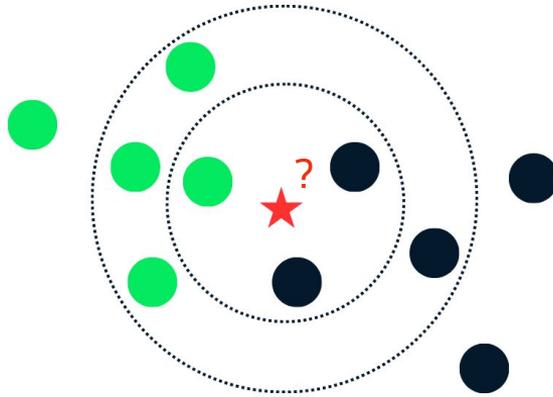
Régression

# Quelques algorithmes de Machine Learning

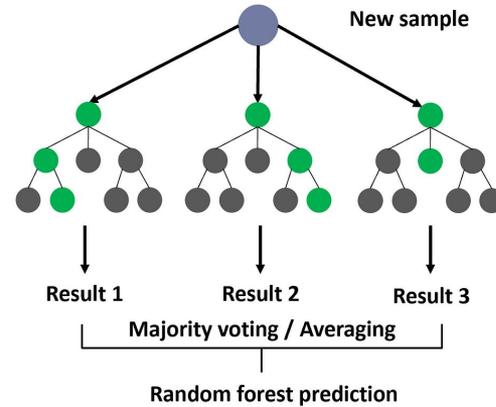


KNN (K plus proche voisin)

# Quelques algorithmes de Machine Learning

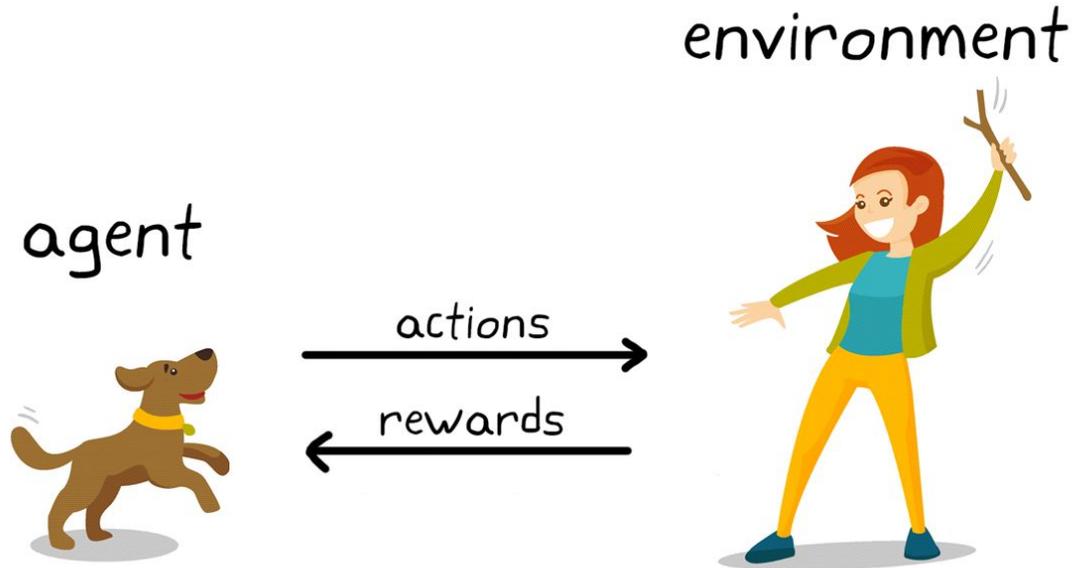


KNN (K plus proche voisin)



Random forest

# Apprentissage par renforcement (Reinforcement Learning)



# Apprentissage par renforcement (Reinforcement Learning)



# Apprentissage par renforcement (Reinforcement Learning)



<https://www.youtube.com/watch?v=WXuK6gekU1Y>



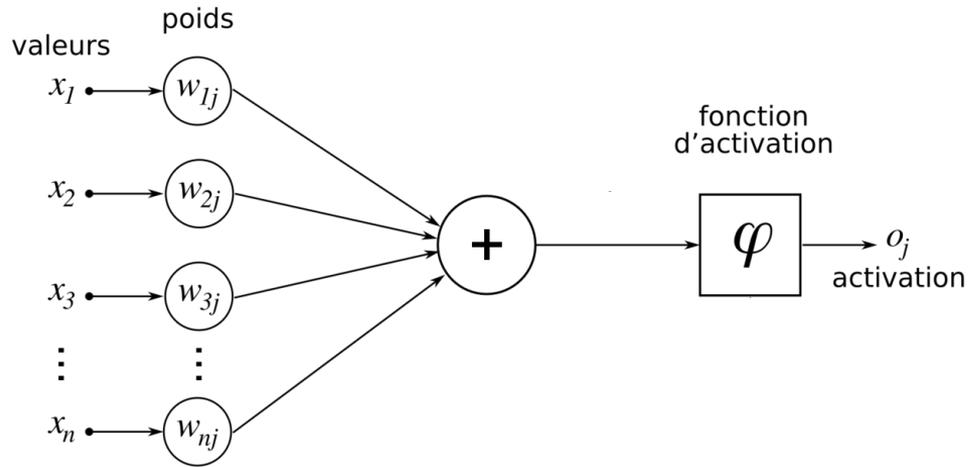
<https://www.youtube.com/watch?v=cUTMhmVh1qs>

# 02

## Comment fonctionne un réseau de neurones artificiels ?

# Un neurone artificiel

> Algorithme Supervisé

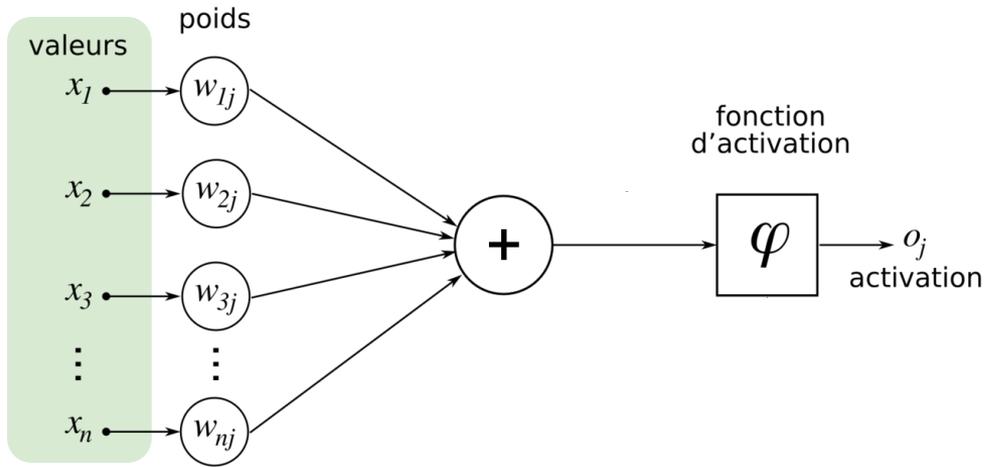


x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

# Un neurone artificiel

> Algorithme Supervisé

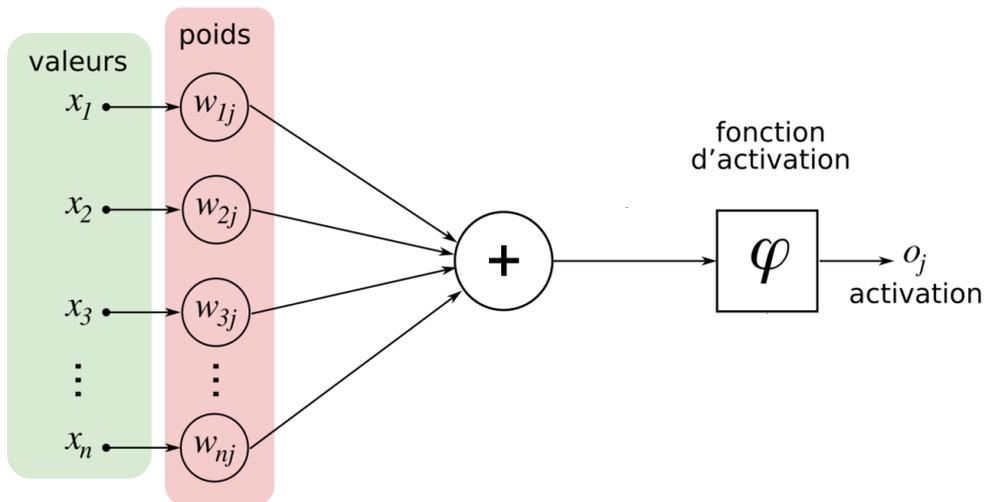
Stimuli



x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

# Un neurone artificiel

Stimuli   Synapses

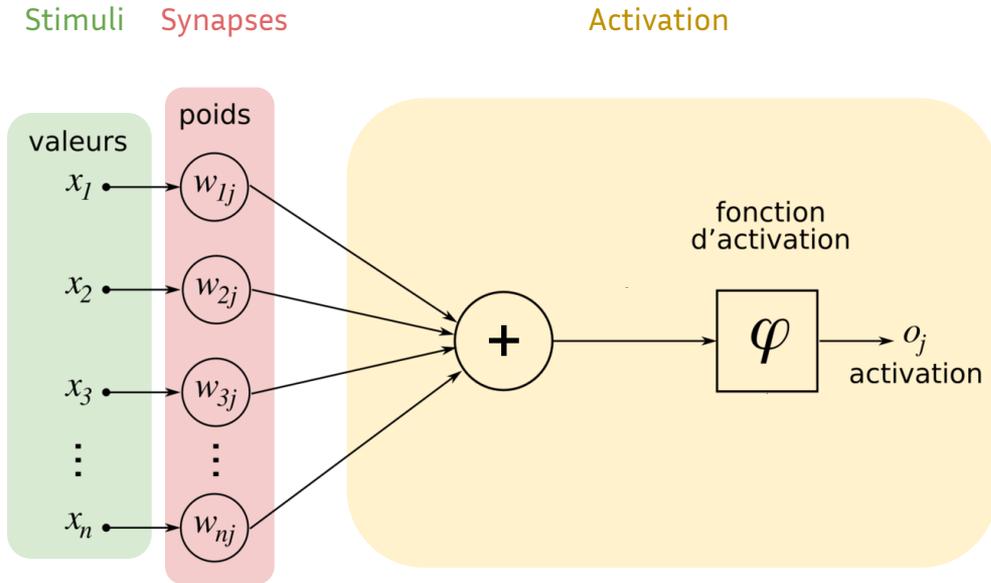


x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

> Algorithme Supervisé

# Un neurone artificiel

> Algorithme Supervisé

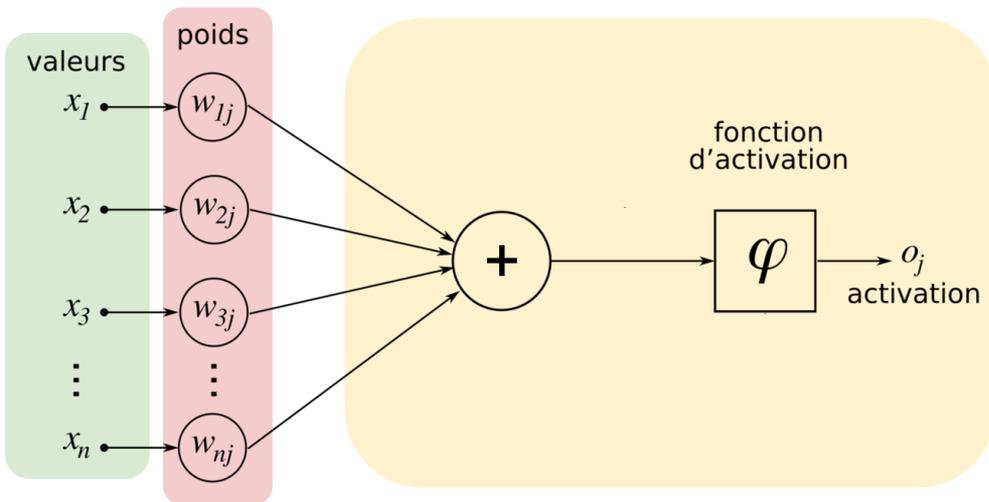


x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

# Un neurone artificiel

Stimuli    Synapses

Activation



x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

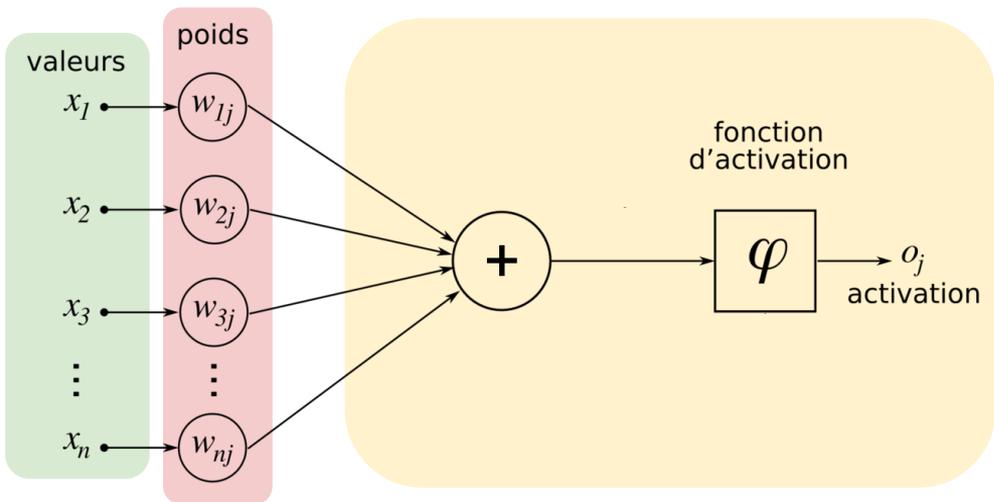
> Algorithme Supervisé

1. On sélectionne un ensemble de poids aléatoires.

# Un neurone artificiel

Stimuli    Synapses

Activation

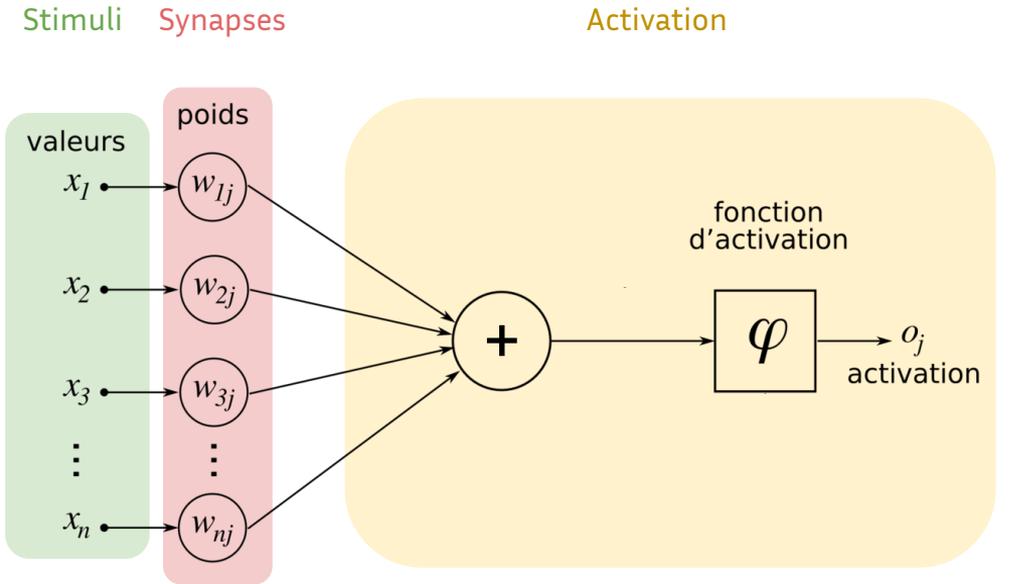


x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

> Algorithme Supervisé

1. On sélectionne un ensemble de poids aléatoires.
2. On fait la prédiction pour J données.

# Un neurone artificiel



x = un nombre  
w = un nombre  
o = un nombre  
e = un nombre

$e_j$   
erreur

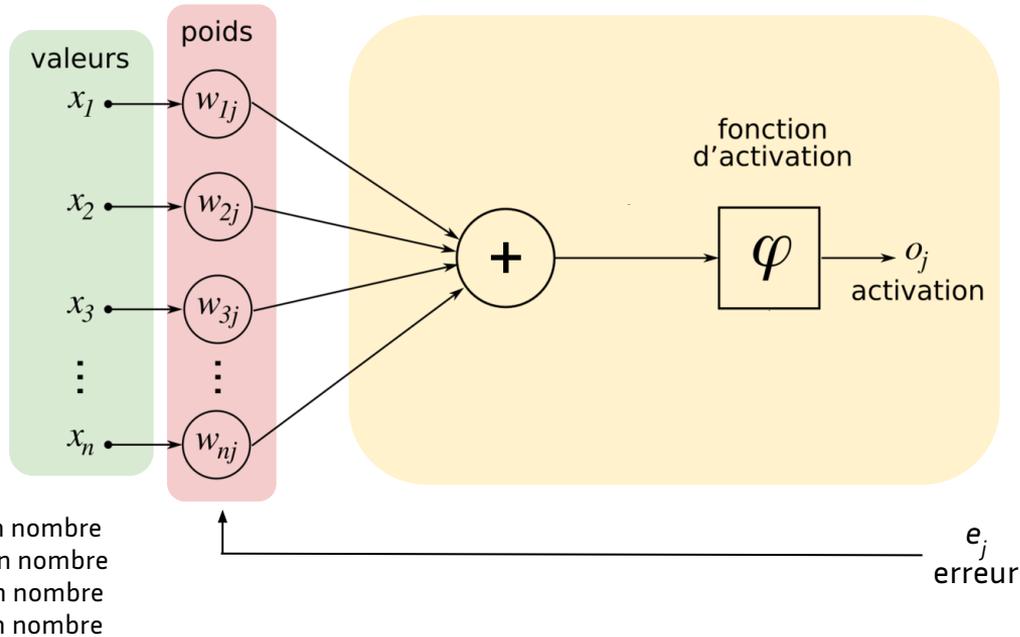
> Algorithme Supervisé

1. On sélectionne un ensemble de poids aléatoires.
2. On fait la prédiction pour J données.
3. On calcul l'erreur moyenne pour ces J données.

# Un neurone artificiel

Stimuli Synapses

Activation



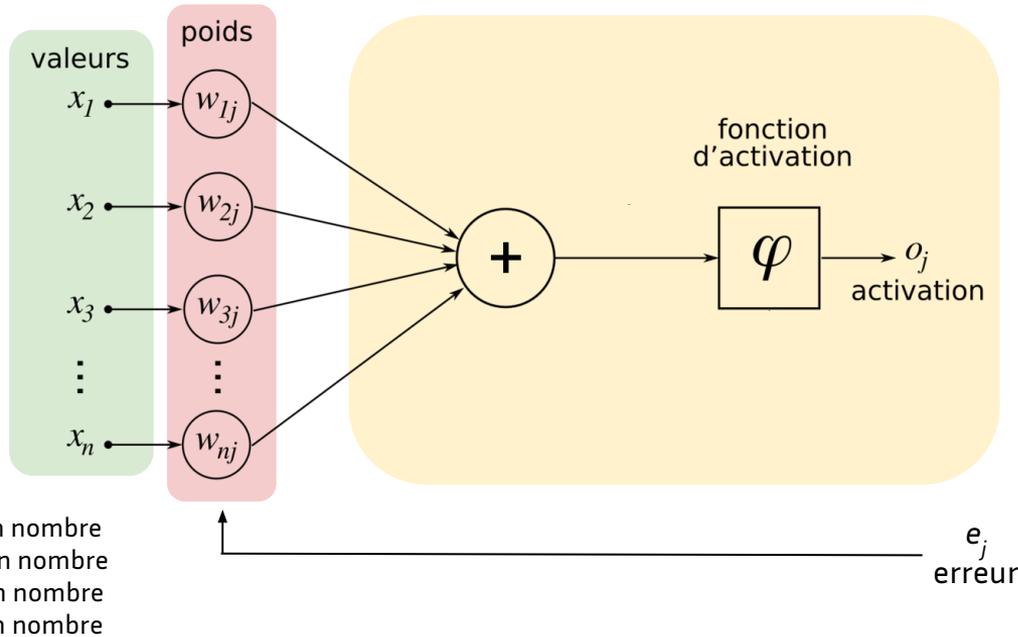
> Algorithme Supervisé

1. On sélectionne un ensemble de poids aléatoires.
2. On fait la prédiction pour J données.
3. On calcul l'erreur moyenne pour ces J données.
4. On modifie nos poids en fonction (descente de gradient).

# Un neurone artificiel

Stimuli Synapses

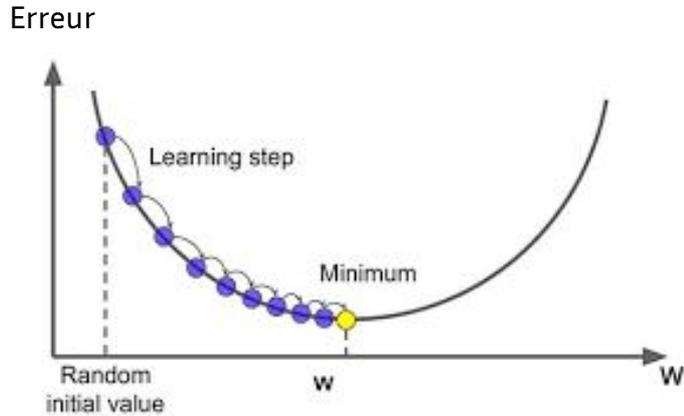
Activation



> Algorithme Supervisé

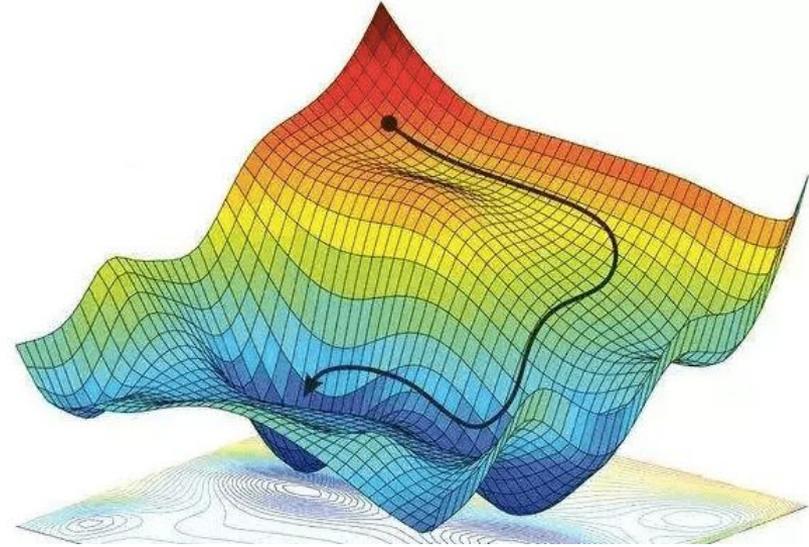
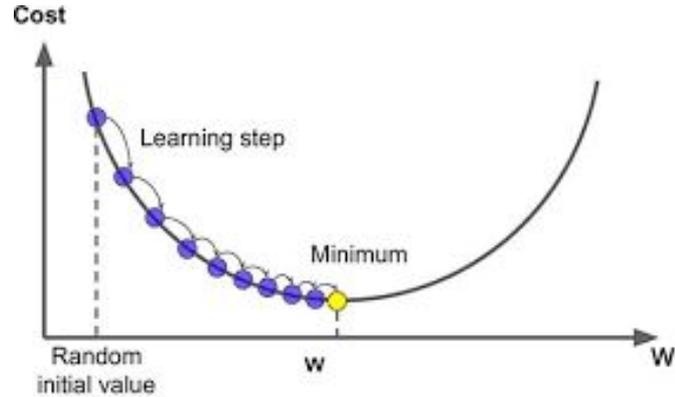
1. On sélectionne un ensemble de poids aléatoires.
2. On fait la prédiction pour J données.
3. On calcul l'erreur moyenne pour ces J données.
4. On modifie nos poids en fonction (descente de gradient).

# Descente du gradient



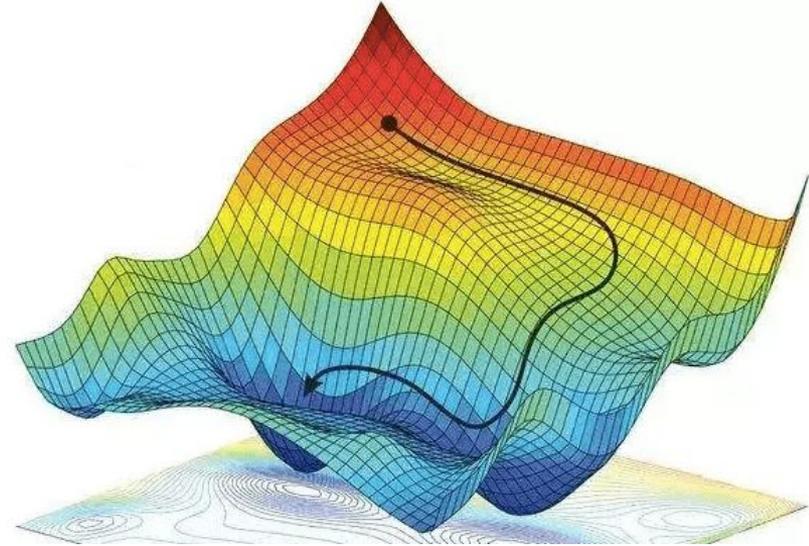
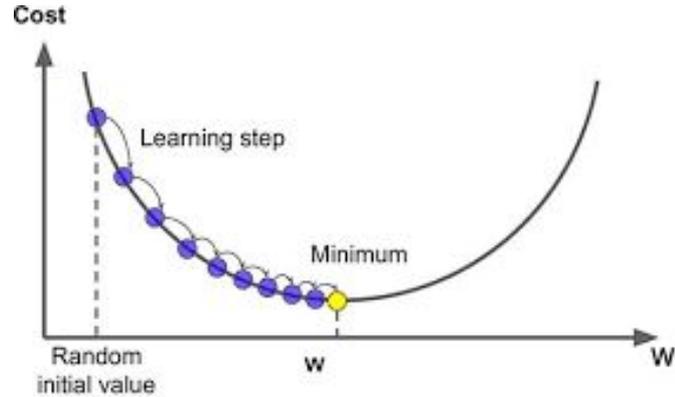
- > Cette fonction représente l'erreur du modèle en fonction du poids  $w$ .
- > On découvre cette fonction petit à petit, au fur et à mesure de l'entraînement.

# Descente du gradient



- > Cette fonction représente l'erreur en fonction du poids  $w$  de notre modèle.
- > On découvre cette fonction petit à petit, au fur et à mesure de l'entraînement.

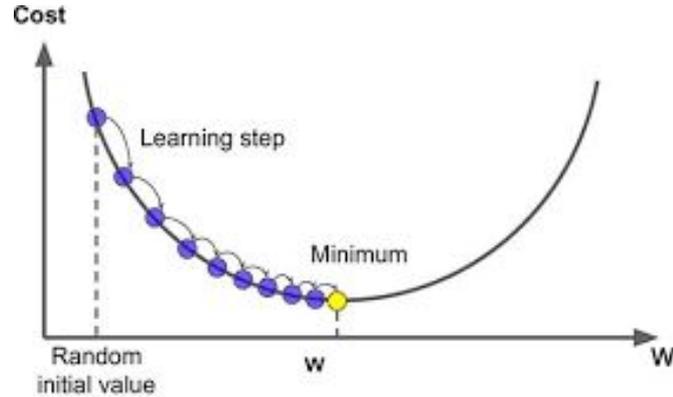
# Descente du gradient



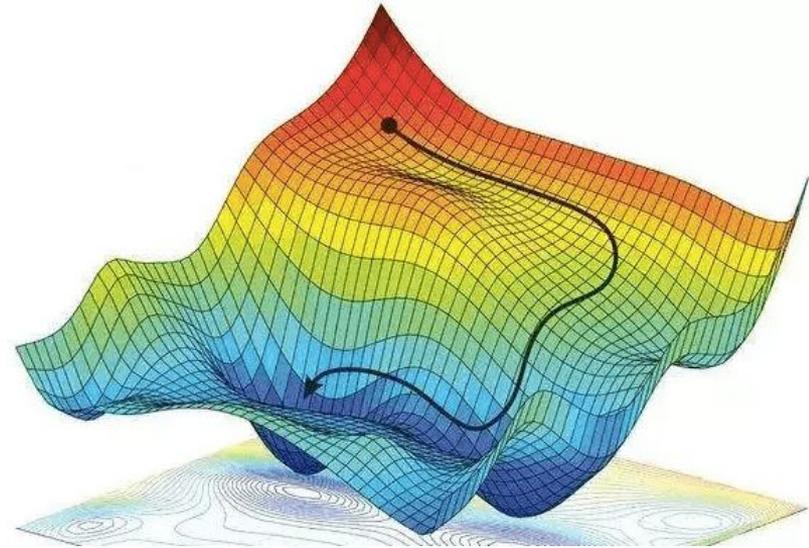
- > Cette fonction représente l'erreur en fonction du poids  $w$  de notre modèle.
- > On découvre cette fonction petit à petit, au fur et à mesure de l'entraînement.



# Descente du gradient



- > Cette fonction représente l'erreur en fonction du poids  $w$  de notre modèle.
- > On découvre cette fonction petit à petit, au fur et à mesure de l'entraînement.

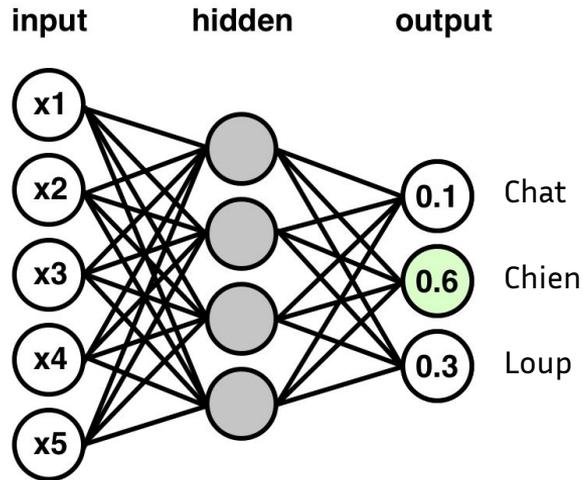


$e$  est l'erreur mesurée en fonction des poids



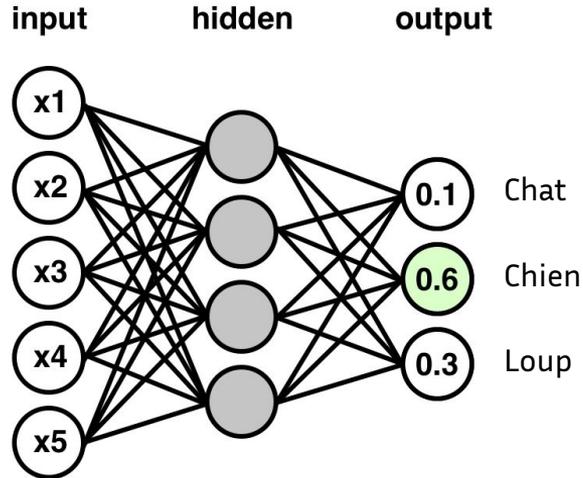
$w_1$  et  $w_2$  sont les poids du modèle

# Comment utiliser un réseau entraîné ?

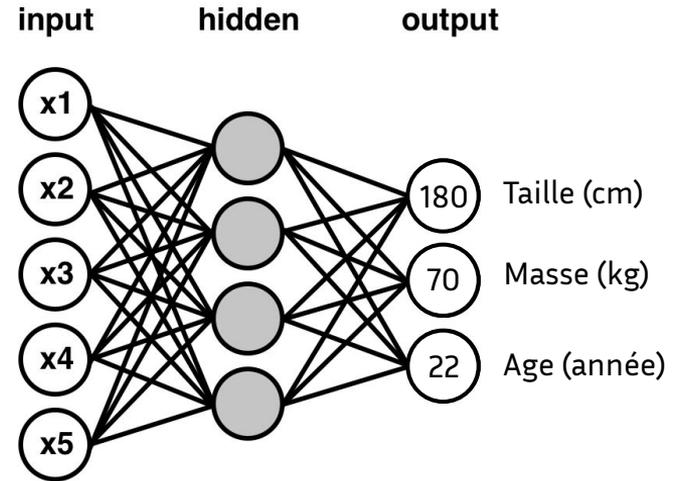


Classification

# Comment utiliser un réseau entraîné ?



Classification



Régression

# Exemple : Word Embeddings

> On transforme des mots en vecteurs de N valeurs.

Queen 

3	3
---	---

King 

1	0
---	---

Woman 

1	4
---	---

Man 

-1	1
----	---

N = 2

# Exemple : Word Embeddings

> On transforme des mots en vecteurs de N valeurs.

Queen	3	3
-------	---	---

King	1	0
------	---	---

N = 2

Woman	1	4
-------	---	---

Man	-1	1
-----	----	---

> On peut faire des opérations arithmétiques avec eux.

$$\text{King} - \text{Queen} = \text{Man} - \text{Woman}$$

$$\text{King} - \text{Queen} + \text{Woman} = \text{Man}$$

# Exemple : Word Embeddings

> On transforme des mots en vecteurs de N valeurs.

Queen	3	3
-------	---	---

King	1	0
------	---	---

N = 2

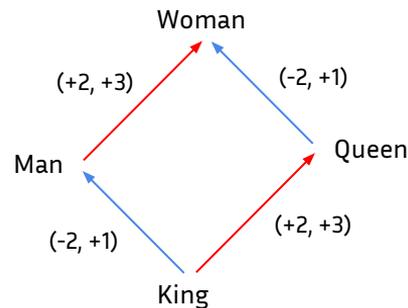
Woman	1	4
-------	---	---

Man	-1	1
-----	----	---

> On peut faire des opérations arithmétiques avec eux.

$$\text{King} - \text{Queen} = \text{Man} - \text{Woman}$$

$$\text{King} - \text{Queen} + \text{Woman} = \text{Man}$$



# 03

## Comment faire avec des séquences ?

# Données séquentielles

> Toute forme de données temporelles :

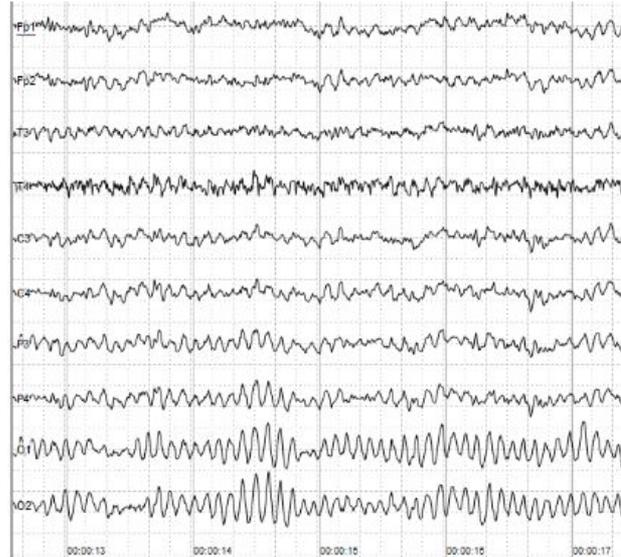
Vidéo, musique, parole, texte, météo, bourse, capteurs, EEG, MEG, fMRI, etc

# Données séquentielles

> Toute forme de données temporelles :

Vidéo, musique, parole, texte, météo, bourse, capteurs, EEG, MEG, fMRI, etc

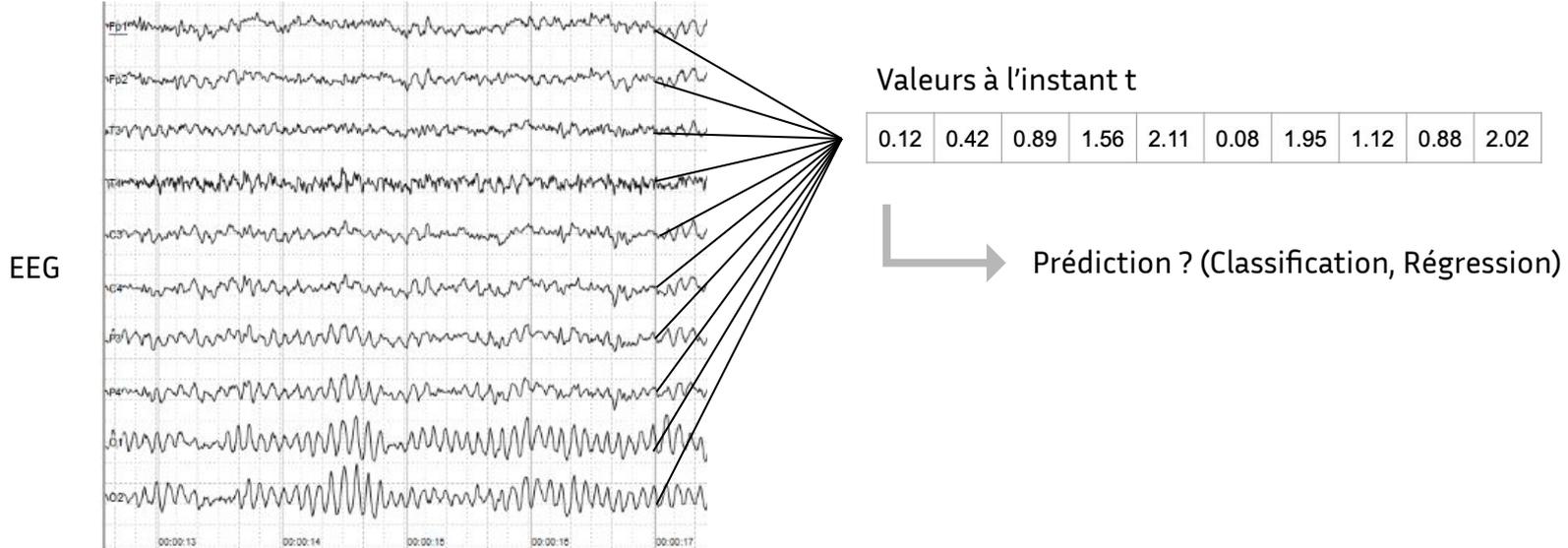
EEG



# Données séquentielles

> Toute forme de données temporelles :

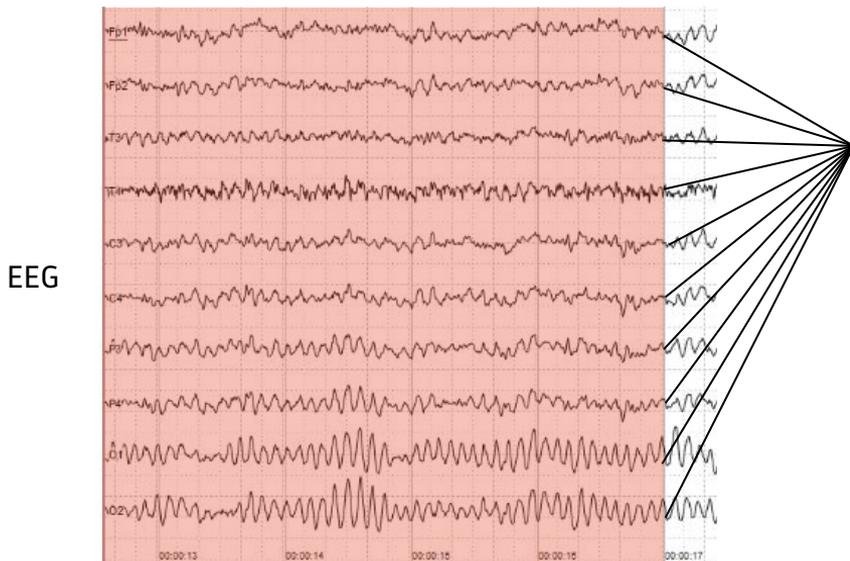
Vidéo, musique, parole, texte, météo, bourse, capteurs, EEG, MEG, fMRI, etc



# Données séquentielles

> Toute forme de données temporelles :

Vidéo, musique, parole, texte, météo, bourse, capteurs, EEG, MEG, fMRI, etc



Valeurs à l'instant t

0.12	0.42	0.89	1.56	2.11	0.08	1.95	1.12	0.88	2.02
------	------	------	------	------	------	------	------	------	------



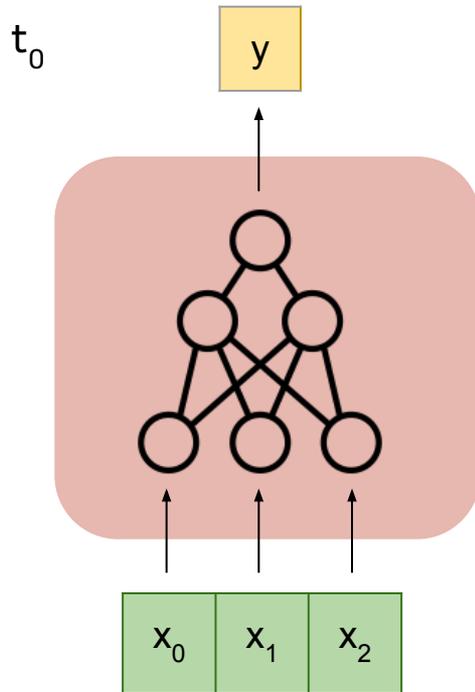
Prédiction ? (Classification, Régression)

Peut-être avons nous besoin du contexte :  
informations des instants précédents (t-1, t-2, t-3, etc)

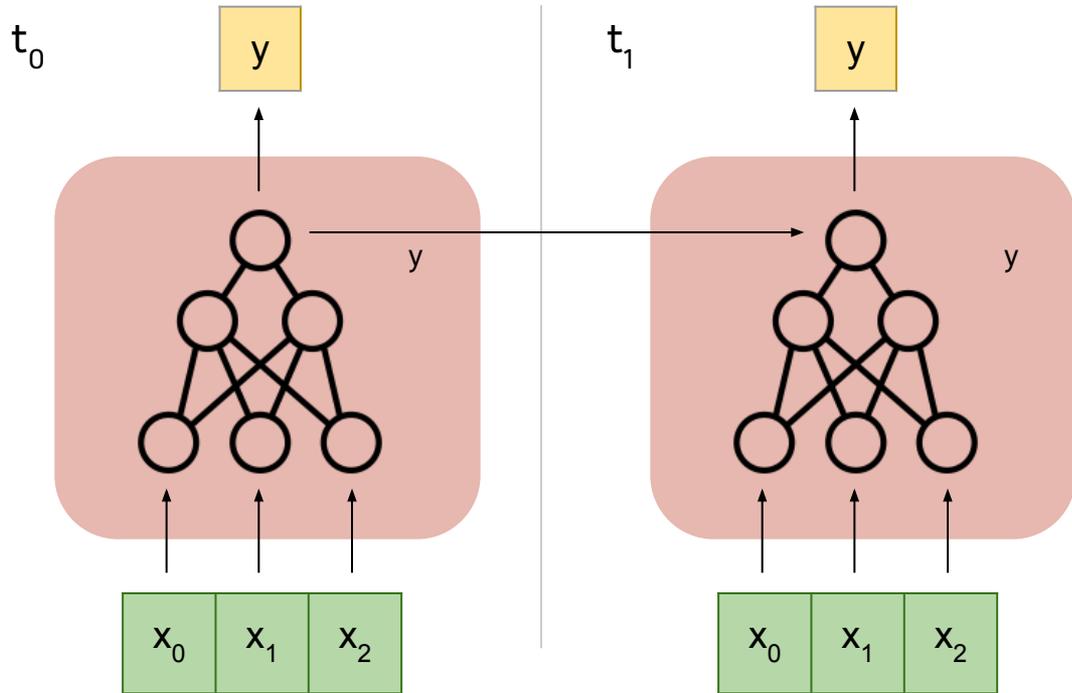
> 1 seconde à 256Hz => 2560 entrées

=> Beaucoup d'entrée !

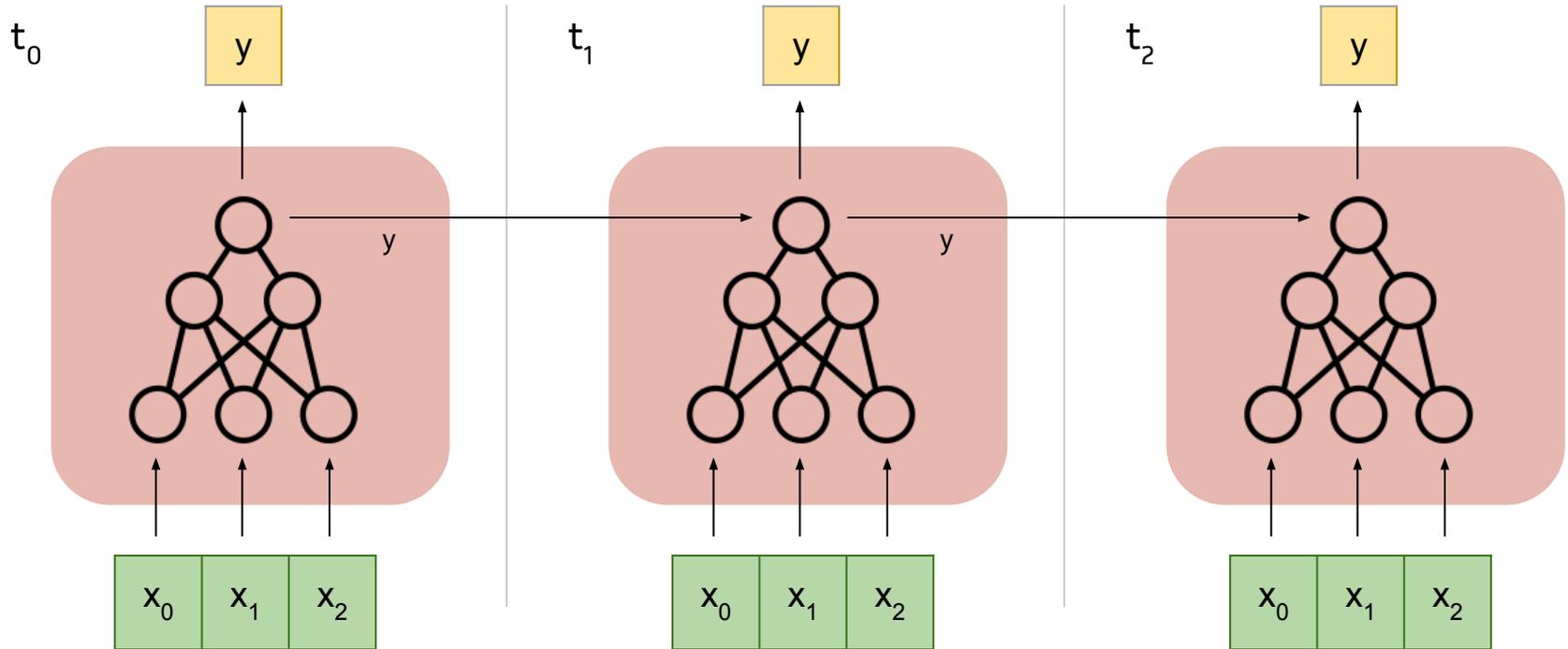
# RNN (Recurrent Neural Network)



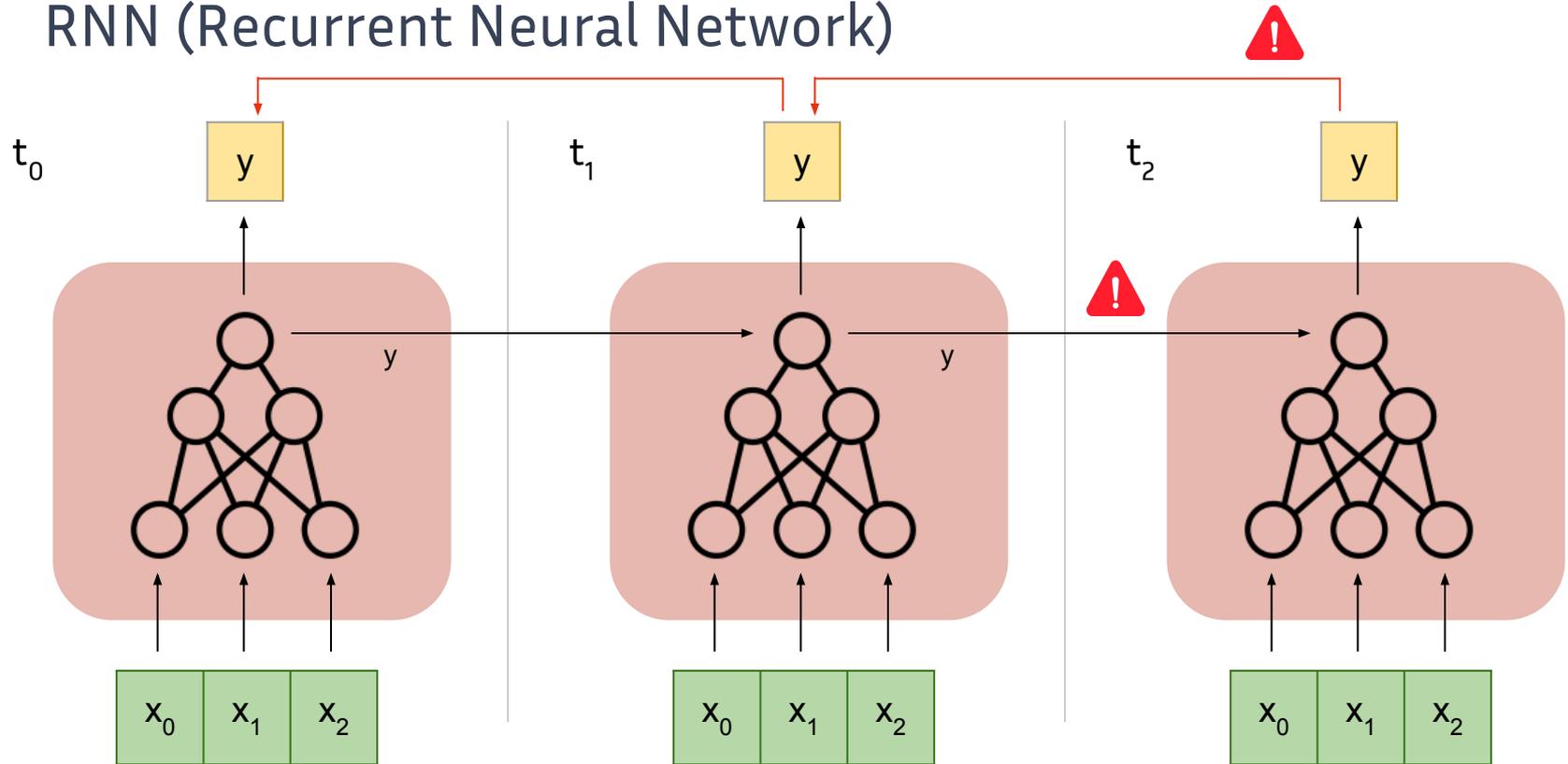
# RNN (Recurrent Neural Network)



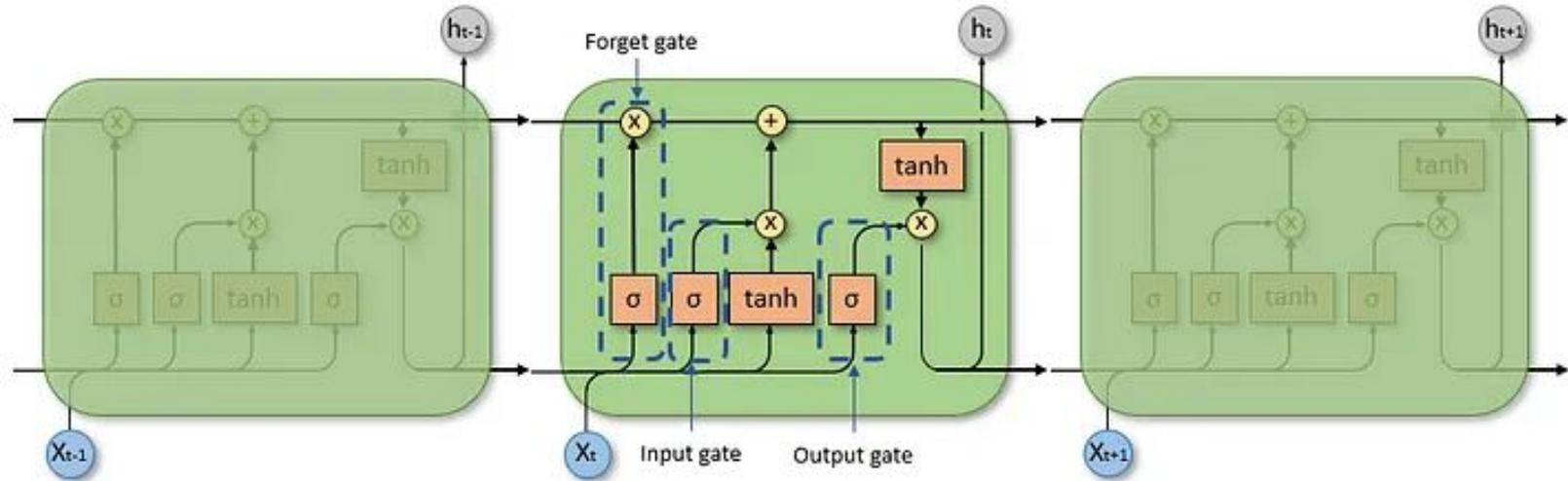
# RNN (Recurrent Neural Network)



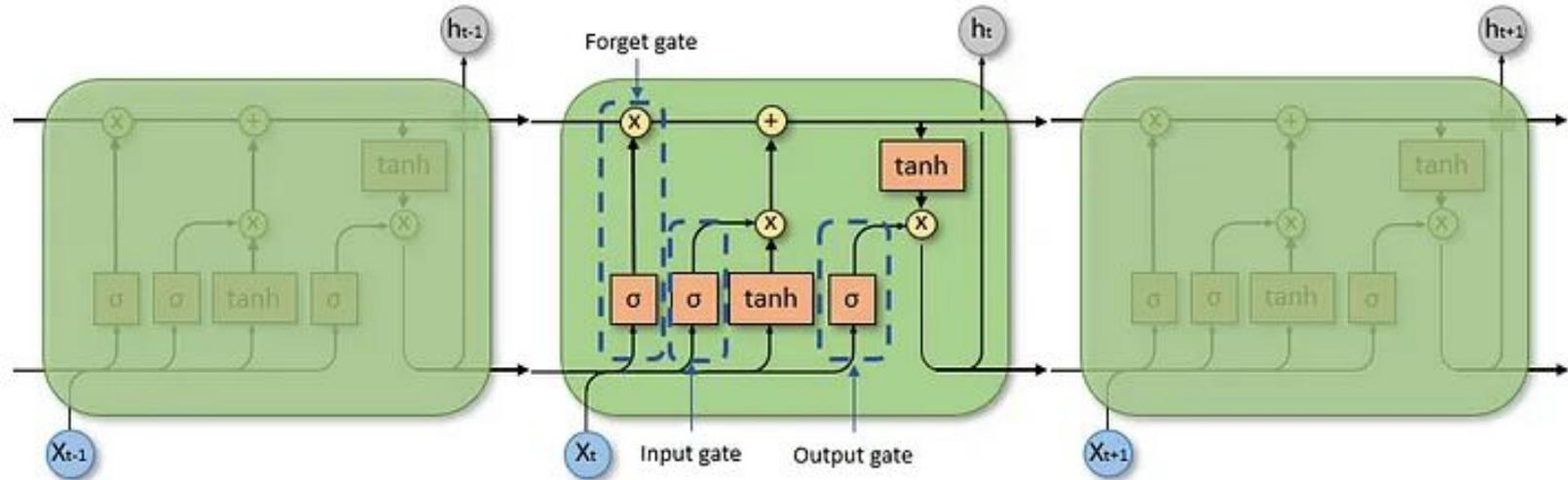
# RNN (Recurrent Neural Network)



# LSTM (Long Short Term Memory)

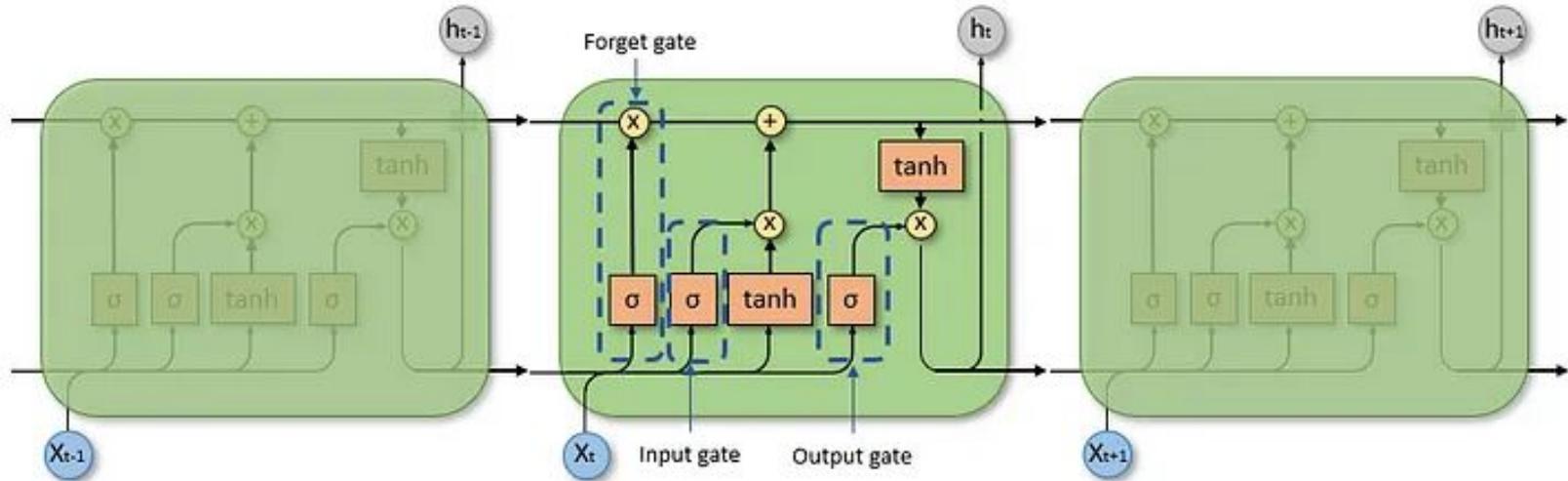


# LSTM (Long Short Term Memory)



- > On transfère la sortie Y
- > On transfère une mémoire C

# LSTM (Long Short Term Memory)



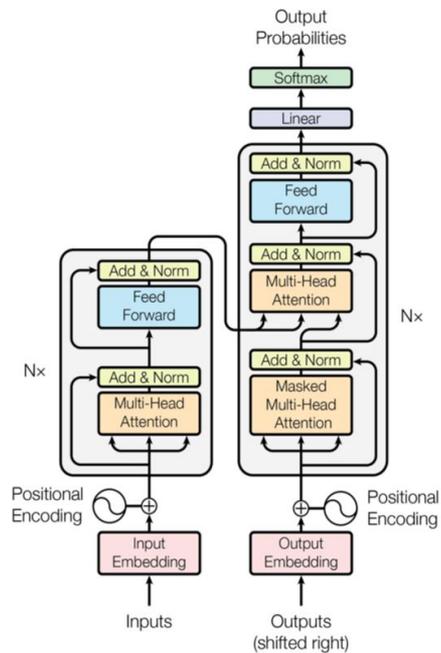
- > On transfère la sortie Y
- > On transfère une mémoire C

- > Grâce à un ensemble de portes, on est capable de mettre à jour la mémoire
- > C'est toujours compliqué de conserver les informations du passé

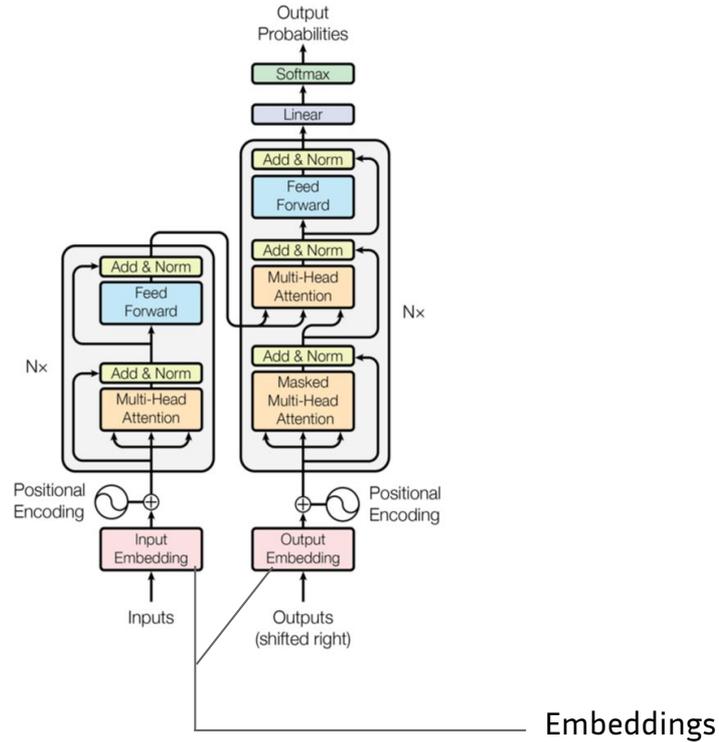
# 04

## Qu'est ce qu'un Transformers ?

# L'architecture d'un Transformers

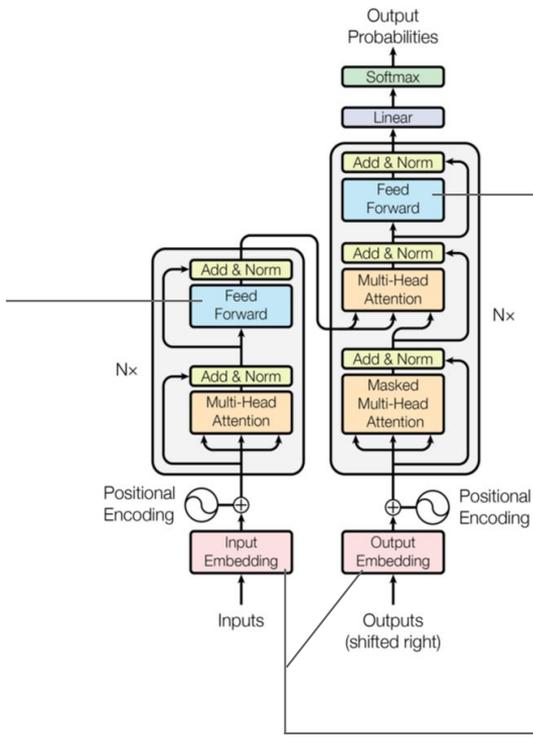


# L'architecture d'un Transformers



# L'architecture d'un Transformers

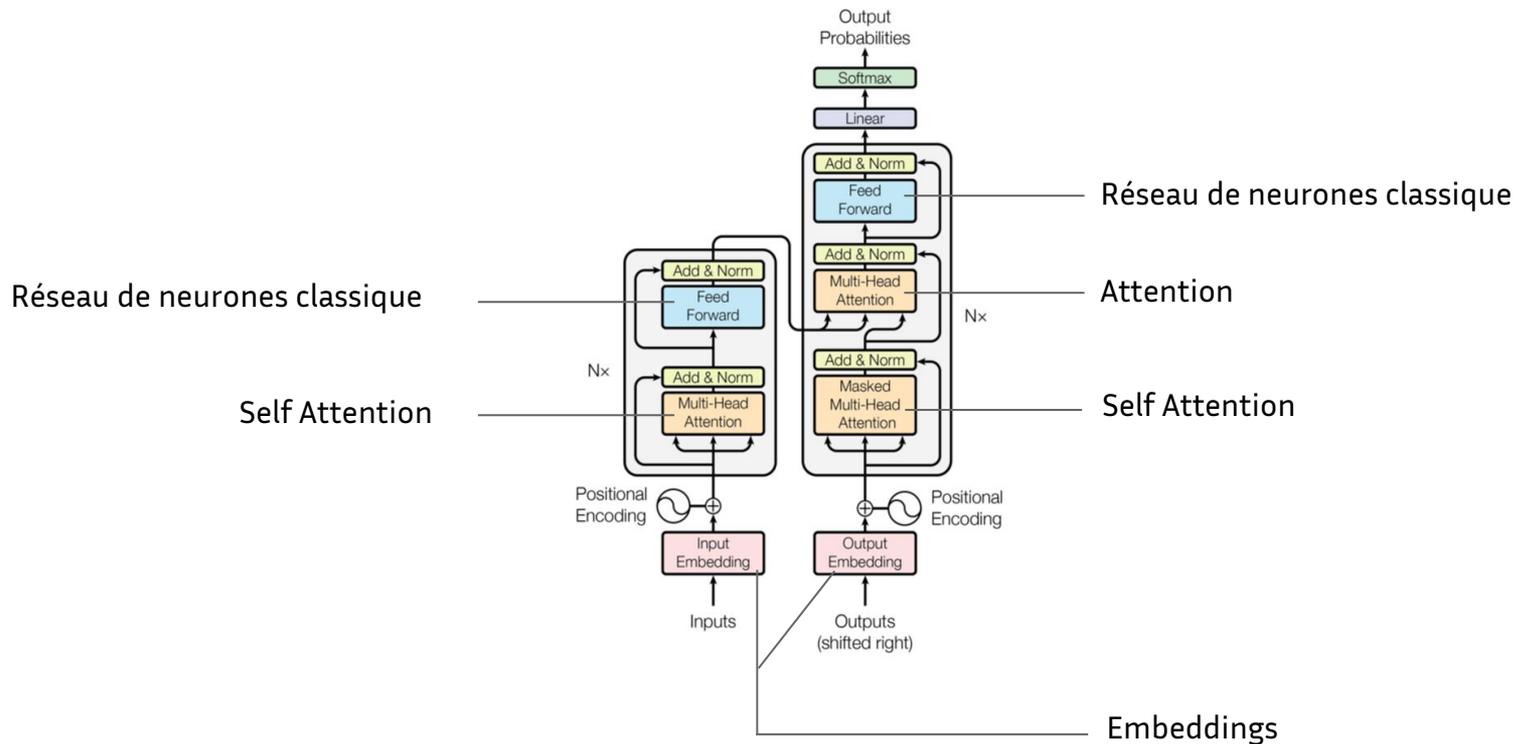
Réseau de neurones classique



Réseau de neurones classique

Embeddings

# L'architecture d'un Transformers



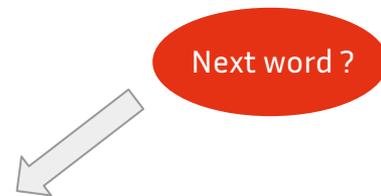
# Qu'est-ce que l'Attention ?

- > On cherche à extraire du passé ( $t-1$ ,  $t-2$ ,  $t-3$ , etc) les informations importantes pour prédire l'instant  $t$ .

# Qu'est-ce que l'Attention ?

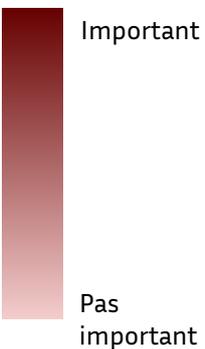
- > On cherche à extraire du passé (t-1, t-2, t-3, etc) les informations importantes pour prédire l'instant t.

"He later went to report Malaysia for one year"



# Qu'est-ce que l'Attention ?

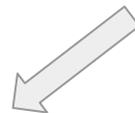
> On cherche à extraire du passé (t-1, t-2, t-3, etc) les informations importantes pour prédire l'instant t.



"He later went to report Malaysia for one year"

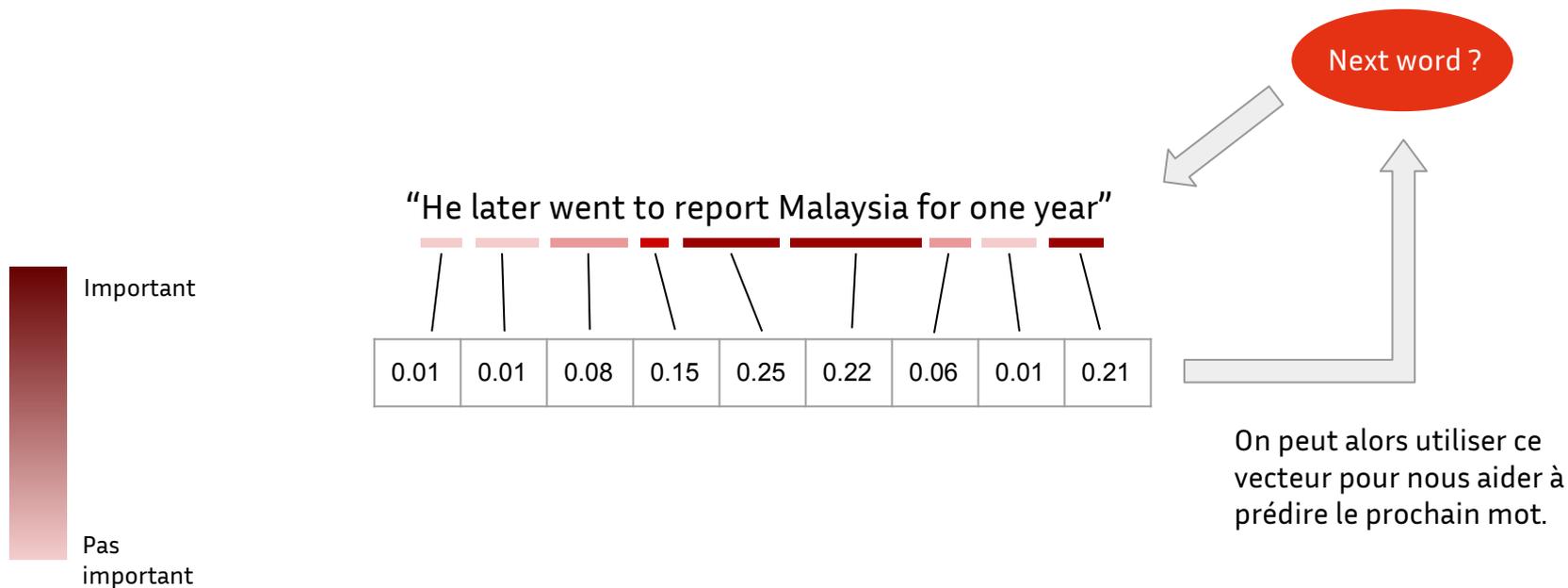


Next word ?



# Qu'est-ce que l'Attention ?

> On cherche à extraire du passé ( $t-1$ ,  $t-2$ ,  $t-3$ , etc) les informations importantes pour prédire l'instant  $t$ .



# Qu'est-ce que l'Attention ?

“He later went to report Malaysia for one year”

# Qu'est-ce que l'Attention ?

“He later went to report Malaysia for one year”

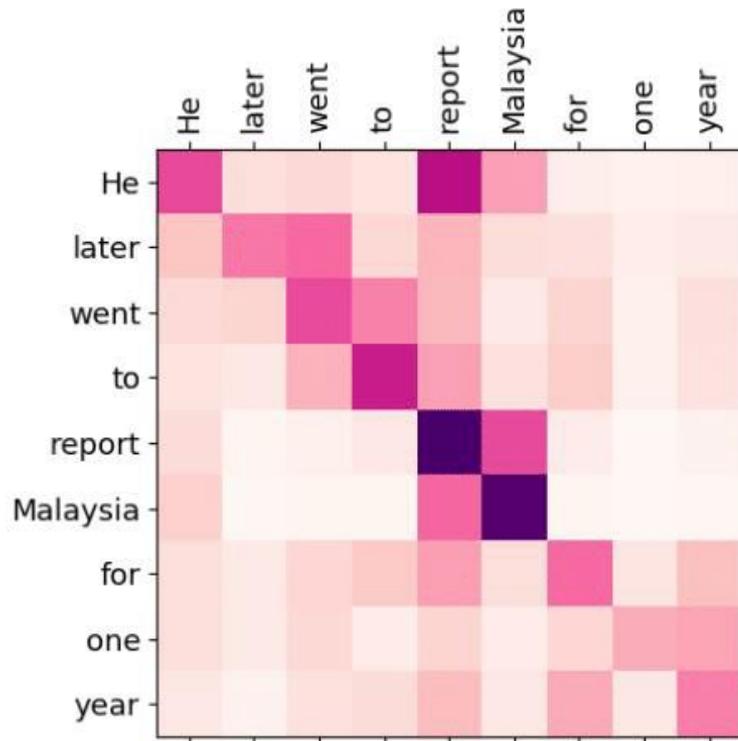
---

> A quel point chaque mot est important pour chaque autre mot ?

# Qu'est-ce que l'Attention ?

“He later went to report Malaysia for one year”

> A quel point chaque mot est important pour chaque autre mot ?

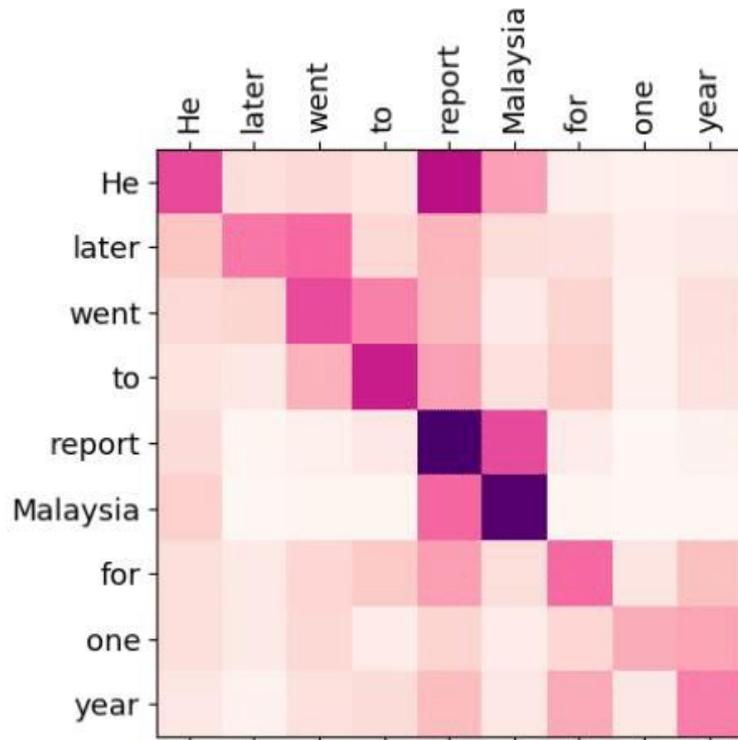


# Qu'est-ce que l'Attention ?

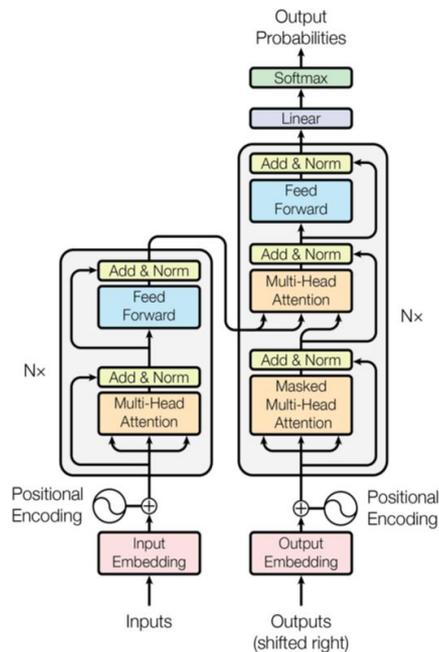
9\*9 = 81 valeurs

“He later went to report Malaysia for one year”

- > A quel point chaque mot est important pour chaque autre mot ?
- > Si notre phrase fait N mots, on obtient alors  $N*N (= N^2)$  différentes valeurs.



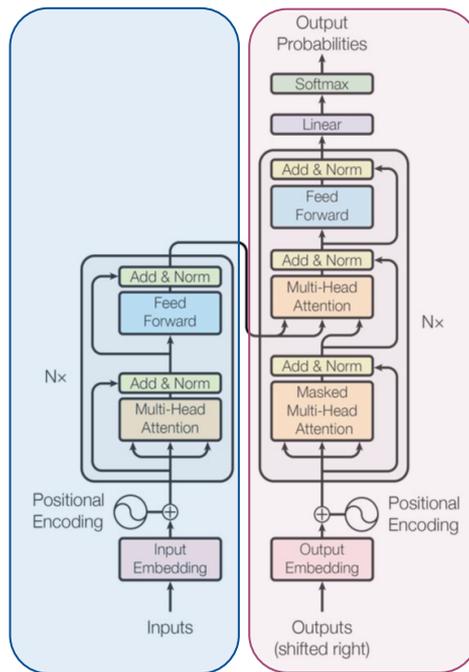
# Encodeur et Décodeur



\* Initialement prévu pour de la traduction

# Encodeur et Décodeur

Encodeur



Décodeur

\* Initialement prévu pour de la traduction

# Encodeur

- > Son but est de compresser les données qu'il reçoit, de manière à ce qu'elle puisse être ensuite décodé par le décodeur.

# Encodeur

- > Son but est de compresser les données qu'il reçoit, de manière à ce qu'elle puisse être ensuite décodée par le décodeur.
- > Il peut être capable de transformer un texte en un simple vecteur (modèle BERT).

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo



0.12	0.42	0.89	1.56	2.11	0.08	1.95	1.12	0.88	2.02
------	------	------	------	------	------	------	------	------	------

# Encodeur

- > Son but est de compresser les données qu'il reçoit, de manière à ce qu'elle puisse être ensuite décodé par le décodeur.
- > Il peut être capable de transformer un texte en un simple vecteur (modèle BERT).

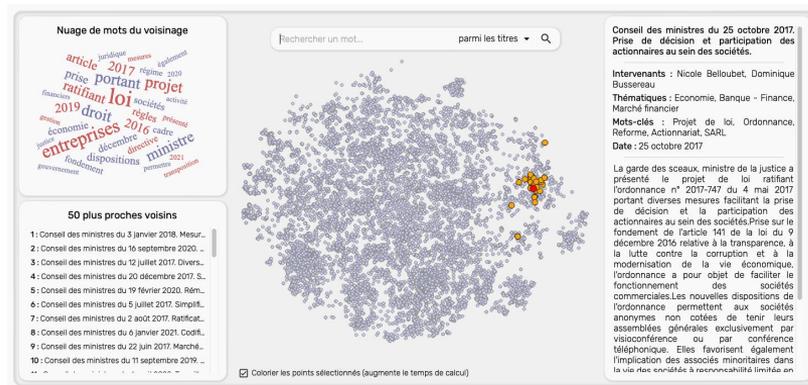
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, non sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo



0.12	0.42	0.89	1.56	2.11	0.08	1.95	1.12	0.88	2.02
------	------	------	------	------	------	------	------	------	------

- > On peut calculer des distances entre des textes.

## Exemple



<https://www.naowak.fr/overton/>

# Décodeur

- > Son but est de décoder des données issues de l'encodeur (initialement traduction).

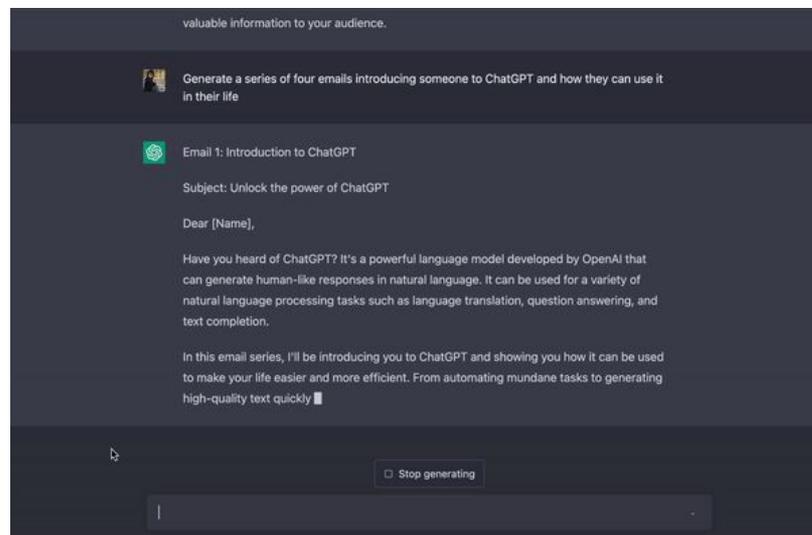
# Décodeur

- > Son but est de décoder des données issues de l'encodeur.
- > On a remarqué que si on ne lui donne pas de données issues de l'encodeur, mais qu'on lui donne le début d'une séquence, il est capable de la continuer.
- > Dans notre cas (texte), il est capable de deviner le prochain mot d'une phrase.

# Décodeur

- > Son but est de décoder des données issues de l'encodeur.
- > On a remarqué que si on ne lui donne pas de données issues de l'encodeur, mais qu'on lui donne le début d'une séquence, il est capable de la continuer.
- > Dans notre cas (texte), il est capable de deviner le prochain mot d'une phrase.

- > C'est chatGPT.



05

# Comment entraîner un transformers ?

# Entraîner un modèle de base

> Il faut choisir une architecture et une taille (nombre de paramètres)

Modèle	Architecture	Nombre paramètres
BERT (2018)	Encodeur	340.000.000 (340M)
GPT-3 (2020)	Décodeur	175.000.000.000 (175B)
GPT-4 (2023)	Décodeur + MOE	1.760.000.000.000 (8 * 220B)

# Entraîner un modèle de base

> Il faut choisir une architecture et une taille (nombre de paramètres)

Modèle	Architecture	Nombre paramètres
BERT (2018)	Encodeur	340.000.000 (340M)
GPT-3 (2020)	Décodeur	175.000.000.000 (175B)
GPT-4 (2023)	Décodeur + MOE	1.760.000.000.000 (8 * 220B)
GPT-3.5-turbo (2023)	Décodeur	20.000.000.000 (20B)
Mistral (2023)	Décodeur	7.000.000.000 (7B)

# Entraîner un modèle de base

> Il faut choisir une architecture et une taille (nombre de paramètres)

Modèle	Architecture	Nombre paramètres
BERT (2018)	Encodeur	340.000.000 (340M)
GPT-3 (2020)	Décodeur	175.000.000.000 (175B)
GPT-4 (2023)	Décodeur + MOE	1.760.000.000.000 (8 * 220B)
GPT-3.5-turbo (2023)	Décodeur	20.000.000.000 (20B)
Mistral (2023)	Décodeur	7.000.000.000 (7B)

> Il faut beaucoup de données (vraiment beaucoup) : tout internet pour chatGPT.

# Entraîner un modèle de base

- > Il faut choisir une architecture et une taille (nombre de paramètres)

Modèle	Architecture	Nombre paramètres
BERT (2018)	Encodeur	340.000.000 (340M)
GPT-3 (2020)	Décodeur	175.000.000.000 (175B)
GPT-4 (2023)	Décodeur + MOE	1.760.000.000.000 (8 * 220B)
GPT-3.5-turbo (2023)	Décodeur	20.000.000.000 (20B)
Mistral (2023)	Décodeur	7.000.000.000 (7B)

- > Il faut beaucoup de données (vraiment beaucoup) : tout internet pour chatGPT.
- > Beaucoup beaucoup de puissance de calcul ! (4 A100 **minimum** pour un modèle de 6B ~ 80.000 euros)

# Entraîner un modèle de base

- > Il faut choisir une architecture et une taille (nombre de paramètres)

Modèle	Architecture	Nombre paramètres
BERT (2018)	Encodeur	340.000.000 (340M)
GPT-3 (2020)	Décodeur	175.000.000.000 (175B)
GPT-4 (2023)	Décodeur + MOE	1.760.000.000.000 (8 * 220B)
GPT-3.5-turbo (2023)	Décodeur	20.000.000.000 (20B)
Mistral (2023)	Décodeur	7.000.000.000 (7B)

- > Il faut beaucoup de données (vraiment beaucoup) : tout internet pour chatGPT.
- > Beaucoup beaucoup de puissance de calcul ! (4 A100 **minimum** pour un modèle de 6B ~ 80.000 euros)
- > **Un énorme paquet de thunes !** 

# Des modèles de base open-source



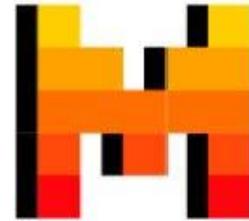
Falcon  
(7B, 40B, 180B)

Emirats Arabes Unis  
Juin 2023



Llama 2  
(7B, 13B, 70B)

Etats-Unis  
Juillet 2023



Mistral  
(7B)

France  
Septembre 2023

# Fine-tuning (optionnel)

- > Un modèle de base ne sait pas dialoguer => deuxième phase d'entraînement

# Fine-tuning (optionnel)

- > Un modèle de base ne sait pas dialoguer => deuxième phase d'entraînement
- > Il faut lui apprendre un format de discussion

system

user

assistant

user

assistant

...

# Fine-tuning (optionnel)

- > Un modèle de base ne sait pas dialoguer => deuxième phase d'entraînement
- > Il faut lui apprendre un format de discussion

**system**

**user**

**assistant**

**user**

**assistant**

...

""

Tu es un assistant, tu dois répondre aux questions que l'utilisateur te pose.

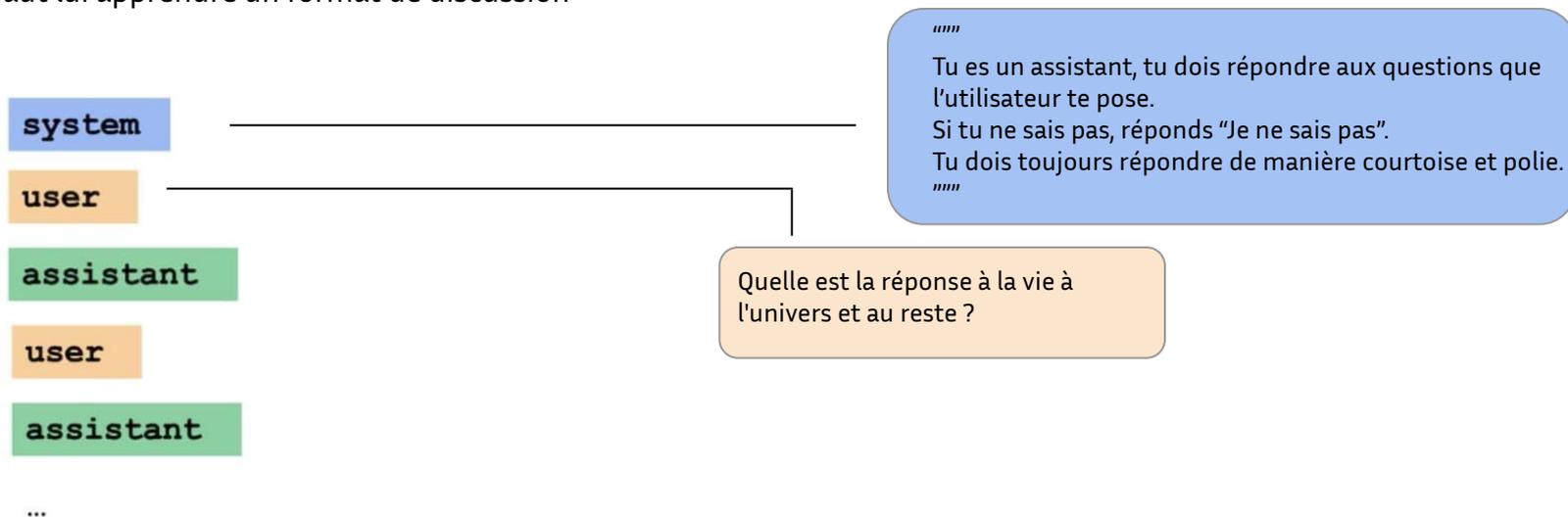
Si tu ne sais pas, réponds "Je ne sais pas".

Tu dois toujours répondre de manière courtoise et polie.

""

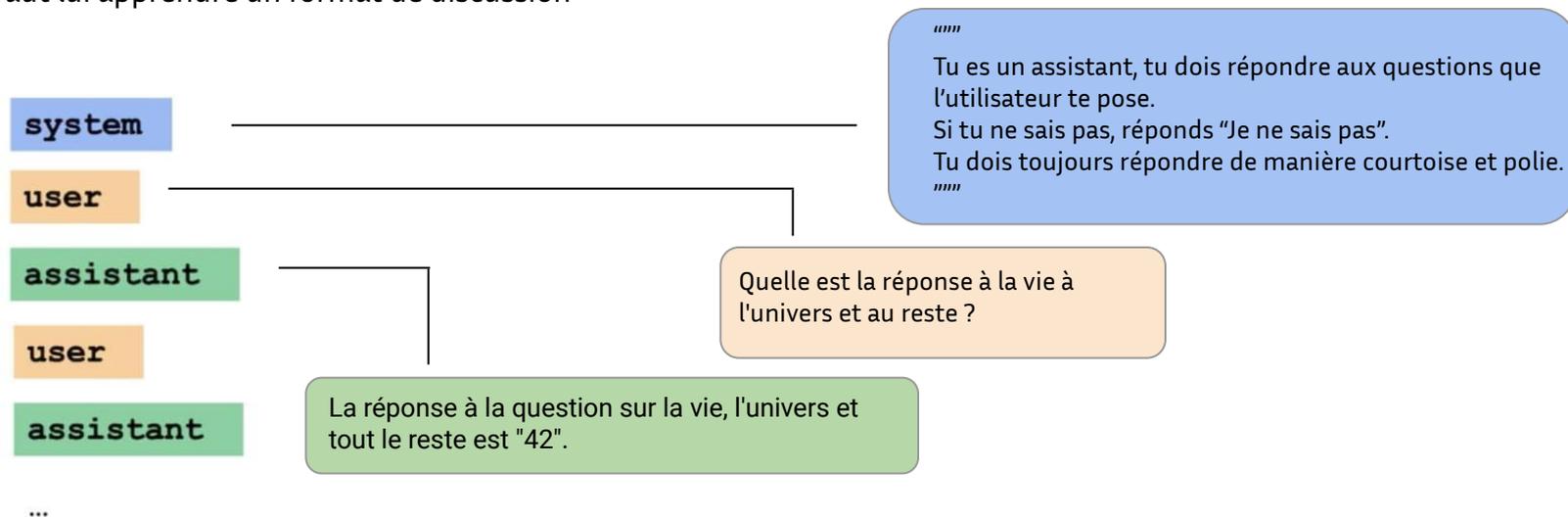
# Fine-tuning (optionnel)

- > Un modèle de base ne sait pas dialoguer => deuxième phase d'entraînement
- > Il faut lui apprendre un format de discussion



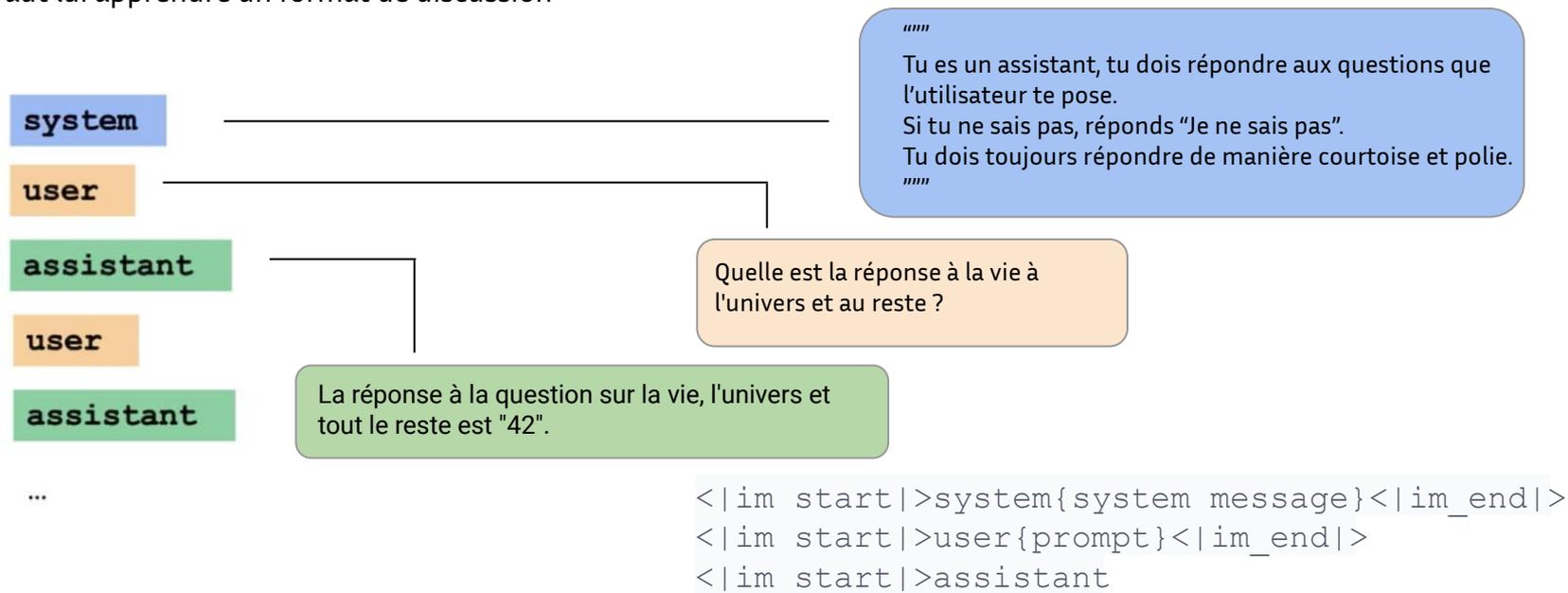
# Fine-tuning (optionnel)

- > Un modèle de base ne sait pas dialoguer => deuxième phase d'entraînement
- > Il faut lui apprendre un format de discussion



# Fine-tuning (optionnel)

- > Un modèle de base ne sait pas dialoguer => deuxième phase d'entraînement
- > Il faut lui apprendre un format de discussion

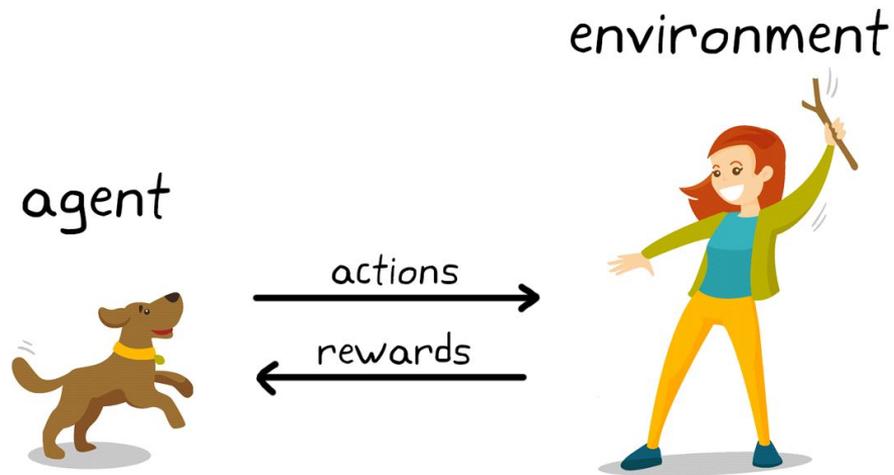


# RLHF (Optionnel)

- > RLHF : Reinforcement Learning with Human Feedback == Apprentissage par Renforcement avec Retour Humain

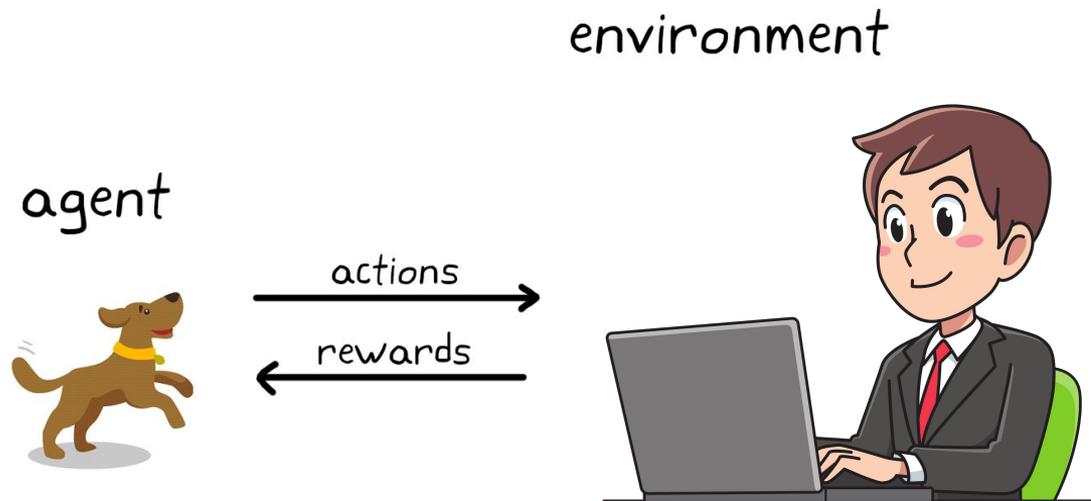
# RLHF (Optionnel)

> RLHF : Reinforcement Learning with Human Feedback == Apprentissage par Renforcement avec Retour Humain



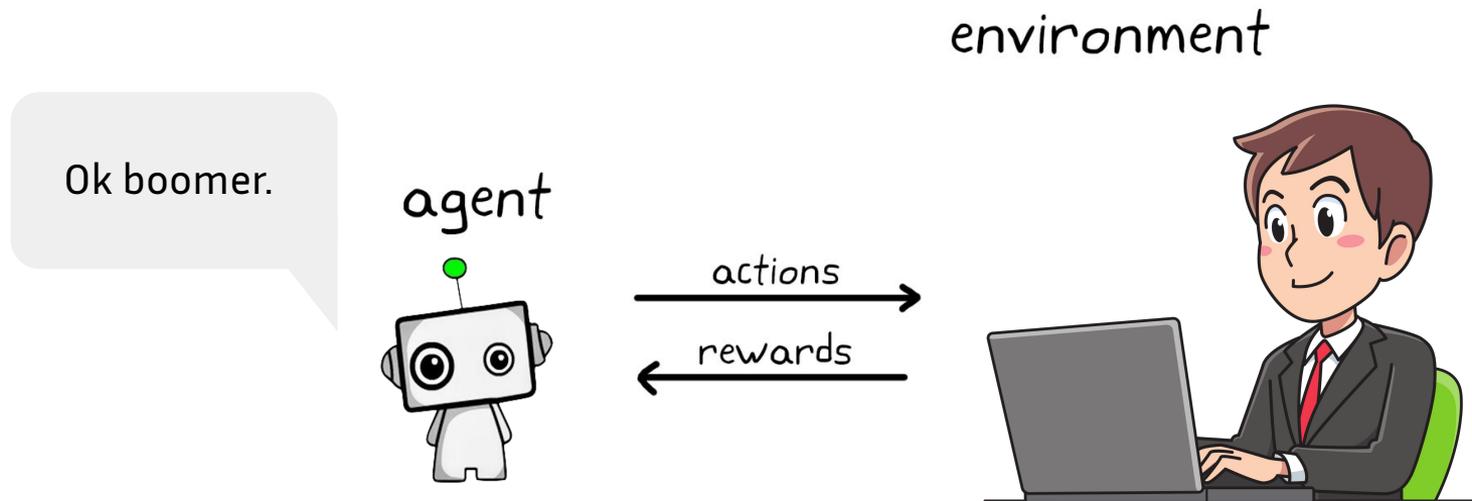
# RLHF (Optionnel)

- > RLHF : Reinforcement Learning with Human Feedback == Apprentissage par Renforcement avec Retour Humain



# RLHF (Optionnel)

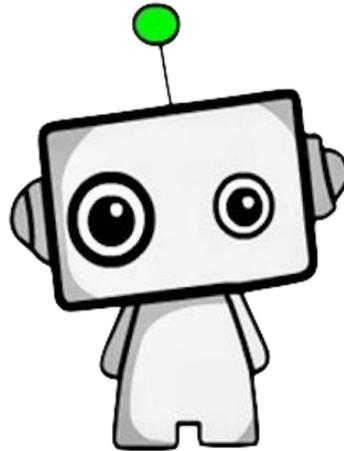
- > RLHF : Reinforcement Learning with Human Feedback == Apprentissage par Renforcement avec Retour Humain



# Prêt pour l'utilisation

> Le modèle est entraîné : on n'y touche plus !

Prêt pour  
discuter !



06

Qu'est ce qu'un prompt ?

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.
- > Mais selon la tâche pour laquelle est entraîné notre modèle, on peut avoir des prompts plus complexe

**system**

**user**

**assistant**

**user**

**assistant**

...

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.
- > Mais selon la tâche pour laquelle est entraîné notre modèle, on peut avoir des prompts plus complexe
- > En fonction du message system, l'ensemble du comportement de notre modèle peut changer : on appelle ça des "agents".

system

user

assistant

user

assistant

...

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.
- > Mais selon la tâche pour laquelle est entraîné notre modèle, on peut avoir des prompts plus complexe
- > En fonction du message system, l'ensemble du comportement de notre modèle peut changer : on appelle ça des "agents".

## Donner une "personnalité"

Auteur, scientifique, humoriste, développeur, cuisiner, etc

system

user

assistant

user

assistant

...

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.
- > Mais selon la tâche pour laquelle est entraîné notre modèle, on peut avoir des prompts plus complexe
- > En fonction du message system, l'ensemble du comportement de notre modèle peut changer : on appelle ça des "agents".

## Donner une "personnalité"

Auteur, scientifique, humoriste, développeur, cuisiner, etc

## Donner des consignes

"Explique étape par étape", "Comme si j'avais 5 ans", "en 1000 mots" etc

system

user

assistant

user

assistant

...

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.
- > Mais selon la tâche pour laquelle est entraîné notre modèle, on peut avoir des prompts plus complexe
- > En fonction du message system, l'ensemble du comportement de notre modèle peut changer : on appelle ça des "agents".

## Donner une "personnalité"

Auteur, scientifique, humoriste, développeur, cuisiner, etc

## Donner des consignes

"Explique étape par étape", "Comme si j'avais 5 ans", "en 1000 mots" etc

## Demander un format de retour

Formatage, code, langue, uniquement un nombre, etc

system

user

assistant

user

assistant

...

# Le prompt : ou comment faire des agents

- > Un début de texte qu'un modèle va compléter/continuer.
- > Mais selon la tâche pour laquelle est entraîné notre modèle, on peut avoir des prompts plus complexe
- > En fonction du message system, l'ensemble du comportement de notre modèle peut changer : on appelle ça des "agents".

## Donner une "personnalité"

Auteur, scientifique, humoriste, développeur, cuisiner, etc

## Donner des consignes

"Explique étape par étape", "Comme si j'avais 5 ans", "en 1000 mots" etc

## Demander un format de retour

Formatage, code, langue, uniquement un nombre, etc

## RAG (Retrieval Augmented Generation)

Donner de l'information / connaissance dans le prompt : très utile pour les Questions/Réponses

system

user

assistant

user

assistant

...

# Des agents dans un environnement



<https://www.naowak.fr/battlefieldagents/>

# Le prompt : tout un futur

- > Il est techniquement possible de “programmer” un comportement pour une IA en langage naturel (fr, en)
- > Le prompt : c’est le moyen de communiquer avec l’IA

# Le prompt : tout un futur

- > Il est techniquement possible de “programmer” un comportement pour une IA en langage naturel (fr, en)
- > Le prompt : c’est le moyen de communiquer avec l’IA
- > Il y a de plus en plus de demande pour des profils de personnes expertes en prompt : Prompt Engineer
- > Les agents conversationnels sont extrêmement efficaces en tant qu’IHM.

# Le prompt : tout un futur

- > Il est techniquement possible de “programmer” un comportement pour une IA en langage naturel (fr, en)
- > Le prompt : c’est le moyen de communiquer avec l’IA
- > Il y a de plus en plus de demande pour des profils de personnes expertes en prompt : Prompt Engineer
- > Les agents conversationnels sont extrêmement efficaces en tant qu’IHM.
- > Des IAs peuvent générer des prompts => Des agents peuvent créer d’autres agents pour résoudre des tâches. (Auto-GPT, dernière version de chatGPT).

# Le prompt : tout un futur

- > Il est techniquement possible de “programmer” un comportement pour une IA en langage naturel (fr, en)
- > Le prompt : c’est le moyen de communiquer avec l’IA
- > Il y a de plus en plus de demande pour des profils de personnes expertes en prompt : Prompt Engineer
- > Les agents conversationnels sont extrêmement efficaces en tant qu’IHM.
- > Des IAs peuvent générer des prompts => Des agents peuvent créer d’autres agents pour résoudre des tâches. (Auto-GPT, dernière version de chatGPT).
- > Pourquoi ça marche ?

Données d'entraînement, inférence statistique, savoir accumulé.

07

Quels sont les risques ?

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

## Différents risques

Données personnelles

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

## Différents risques

Données personnelles

Propagation biais passés

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

## Différents risques

Données personnelles

Propagation biais passés

Centralisation du pouvoir

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

## Différents risques

Données personnelles

Biais induits (propagande, manipulation de masse)

Propagation biais passés

Centralisation du pouvoir

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

## Différents risques

Données personnelles

Propagation biais passés

Centralisation du pouvoir

Biais induits (propagande, manipulation de masse)

Armement : robots tueurs autonomes

# Les risques

- > Beaucoup de risques : mais probablement pas comme dans les films dystopiques
- > L'Homme est le vrai danger

## Différents risques

Données personnelles

Propagation biais passés

Centralisation du pouvoir

Biais induits (propagande, manipulation de masse)

Armement : robots tueurs autonomes

Confiance aveugle

08

# Débat / Questions

# Merci !

yannis.bendi-ouis@inria.fr