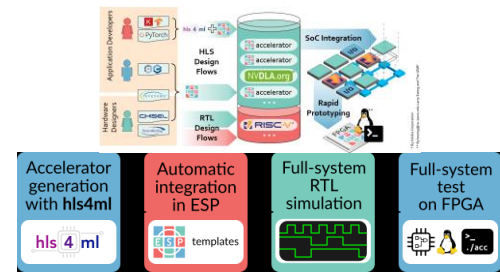
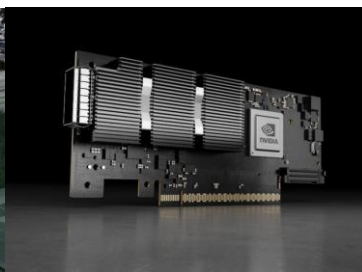
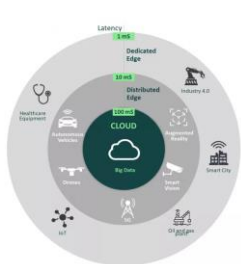


Principes et vocabulaire de l'IA moderne



Approche déductive

Symbolisme

Données d'entrée

ALGORITHMES



Données de sortie

EXPERTS:
LOIS
Règles

Boucles
Tests logiques

Traitement de cas
particuliers:
Logiciels spécialisés

Approche inférentielle (Inductive)

connectivisme

+ Probabilisme

+ Analogisme

Données d'entrée

Données de sortie

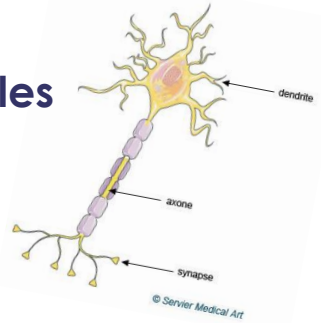


Modèle
d'inférence

MESURES
FAITS
Informations

Algorithmes
apprenants

Lois
Règles



L'intelligence Artificielle 2012 - ...

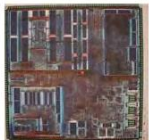
Nouvelles architectures matérielles

Outils de générations de données

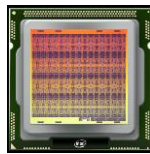
Réseaux de neurones et descente de gradient



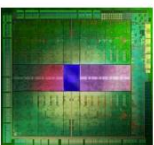
CPUs+ NNproc



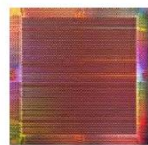
DSPs



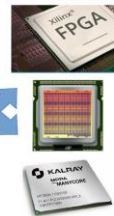
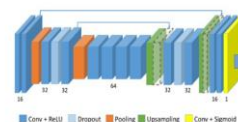
NMC



GPUs



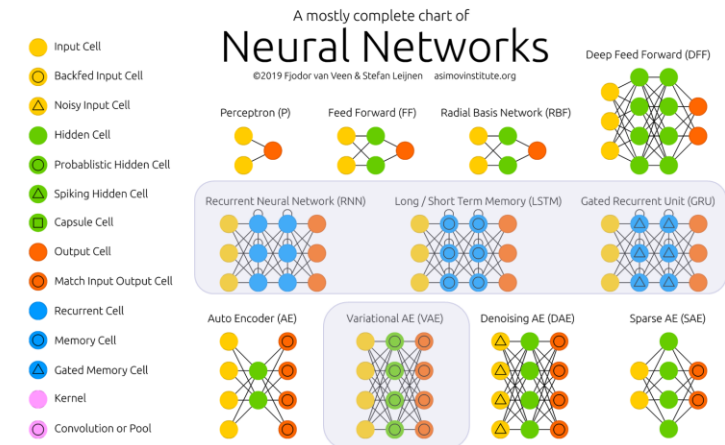
FPGAs

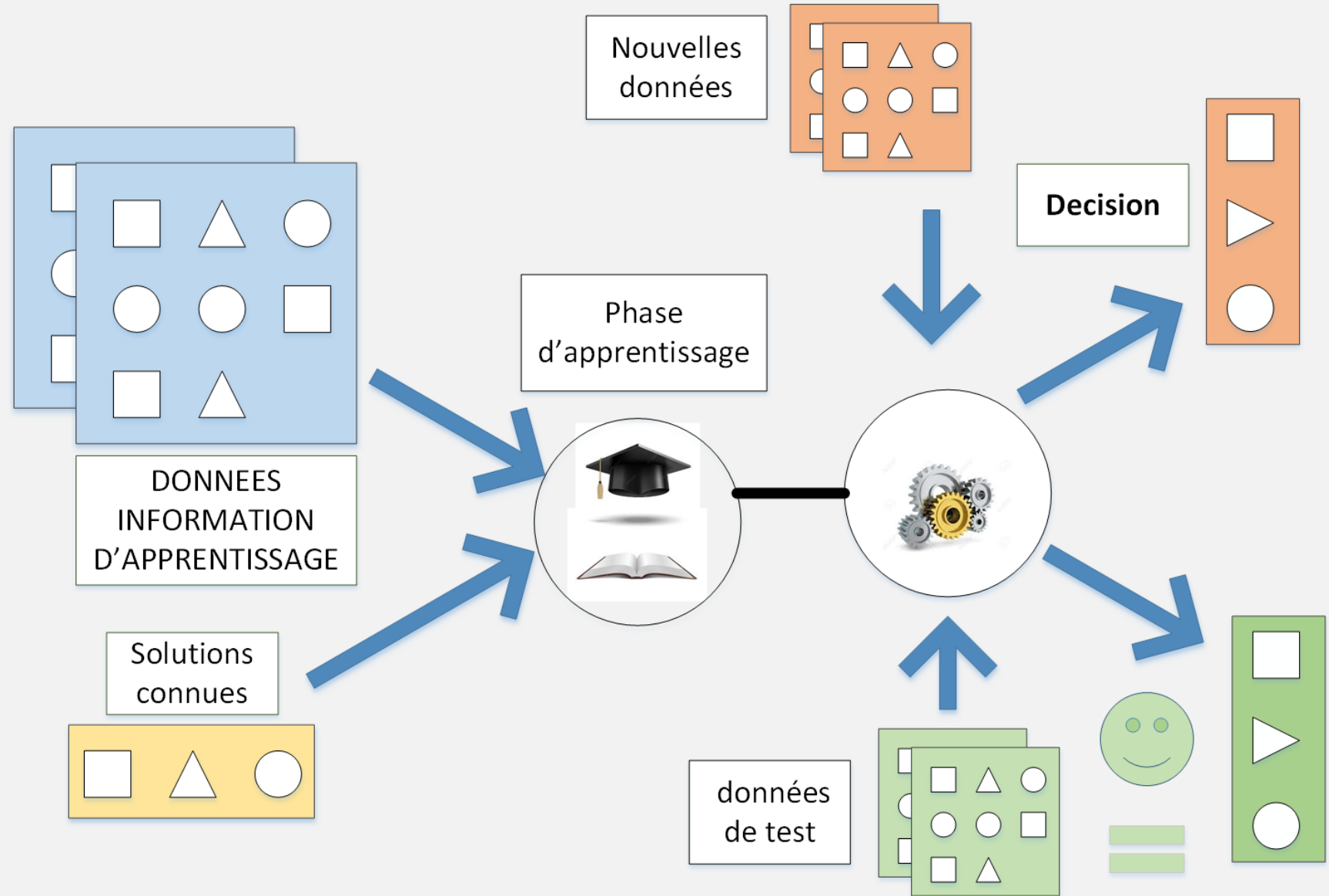


SciPy



AI
Custom Data
and Models
roboflow





Les quatre catégories d'apprentissage profond

LLM

Fine-tuning

Supervised Learning

- Makes machine Learn explicitly
- Data with clearly defined output is given
- Direct feedback is given
- Predicts outcome/future
- Resolves classification and regression problems



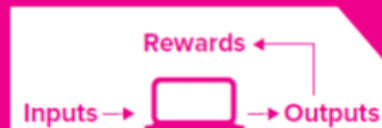
Unsupervised Learning

- Machines understand the data (Identifies patterns/structures)
- Evolution is qualitative or indirect
- Does not predict/find anything specific



Reinforcement Learning

- An approach to AI
- Reward based learning
- Learning from +ve & -ve reinforcement
- Machine learns how to act in a certain environment
- To maximize rewards



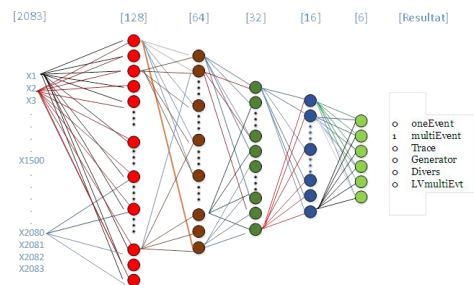
Transfer learning

- Classical supervised techniques
- Change part of the model with new inputs and outputs
- Add features to the preliminary model

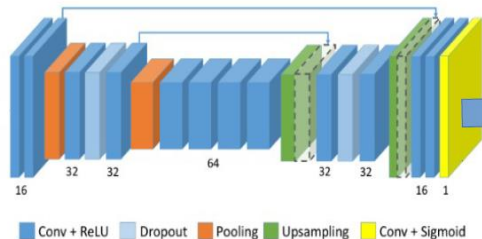


Les briques de bases

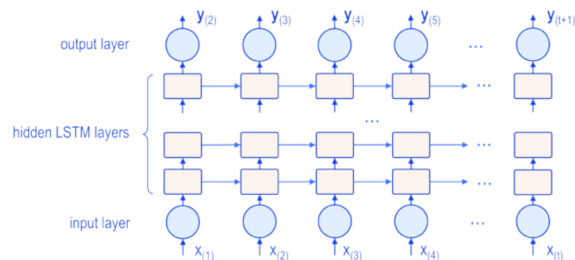
DEEP NEURAL NETWORK



CONVOLUTIONAL NEURAL NETWORK



RECURRENT NEURAL NETWORK



DECISION TREE



RANDOM FOREST



Classique

- Séries temporelles (LLM, ChatGPT)
- Données Audio
- Générateur d'information en apprentissage profond
- Apprentissage par renforcement
- Traitement des données avec graphes (chaînes)
- Traitement des images et des vidéos (computer Vision)
- Traitement du langage naturel
- Structuration des données



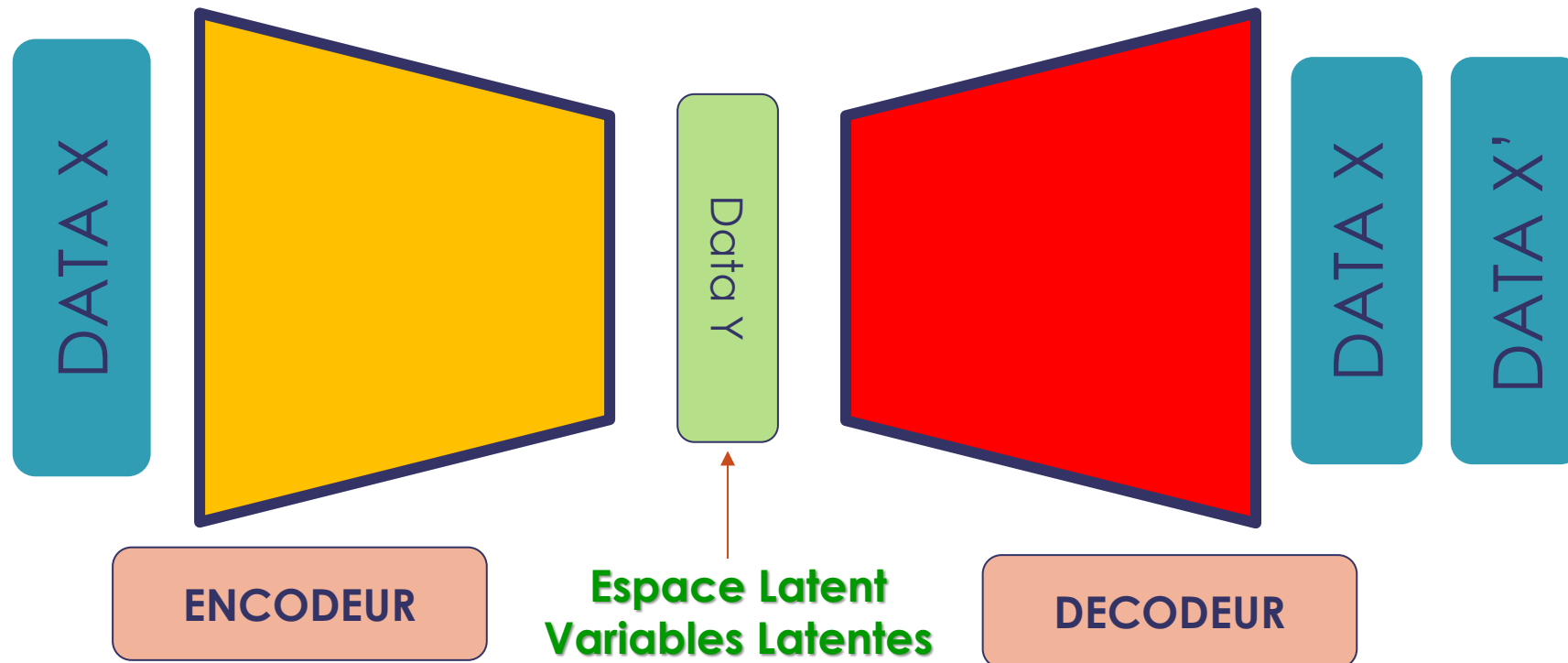
Et si l'IA avait de l'imagination

- Réaliser la labellisation des données est ennuyeux, couteux
- Comment faire pour automatiser cette phase de labellisation ?
- ➔ laisser le modèle découvrir les clusters !

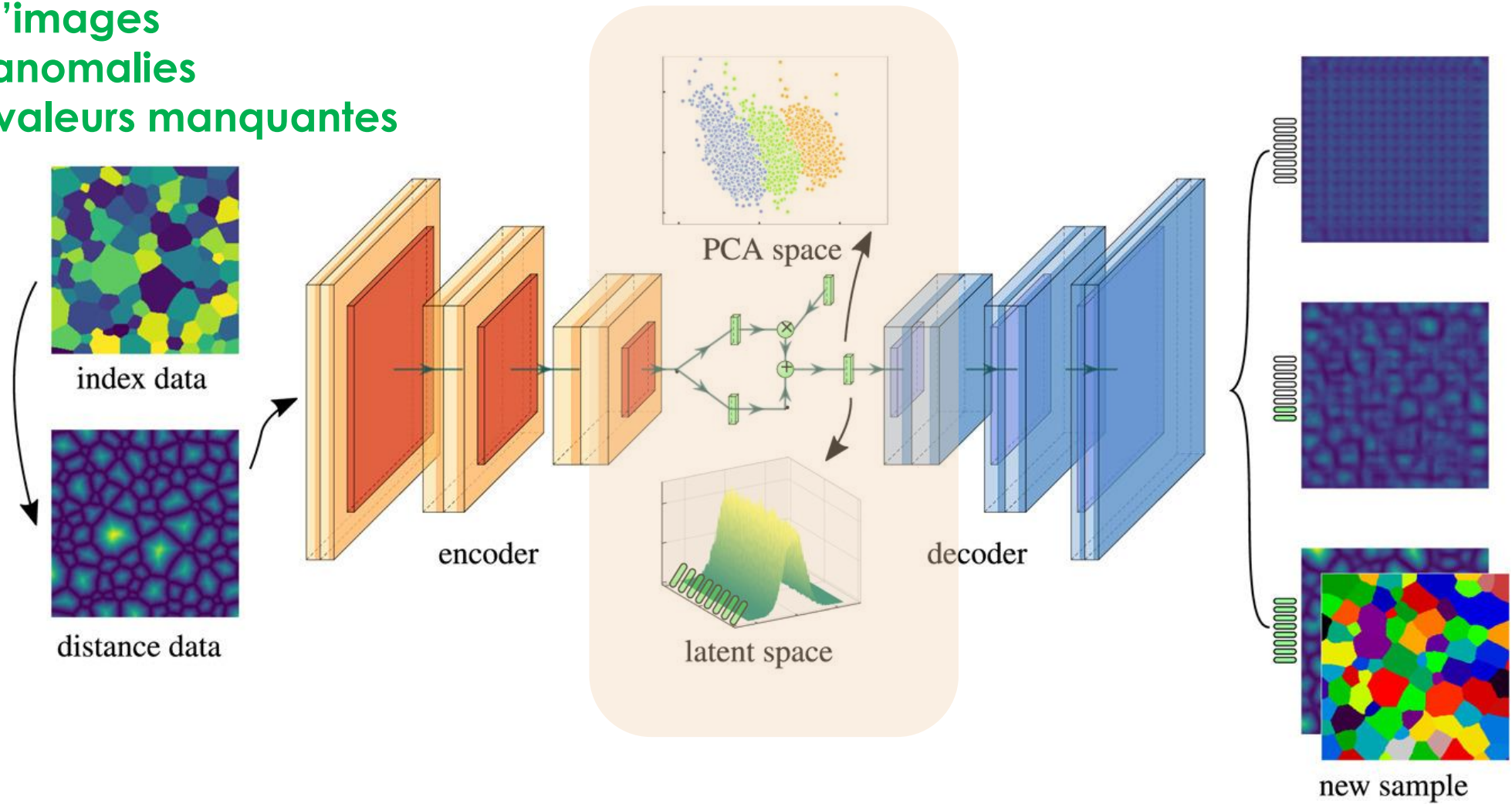
On ne connaît que les X

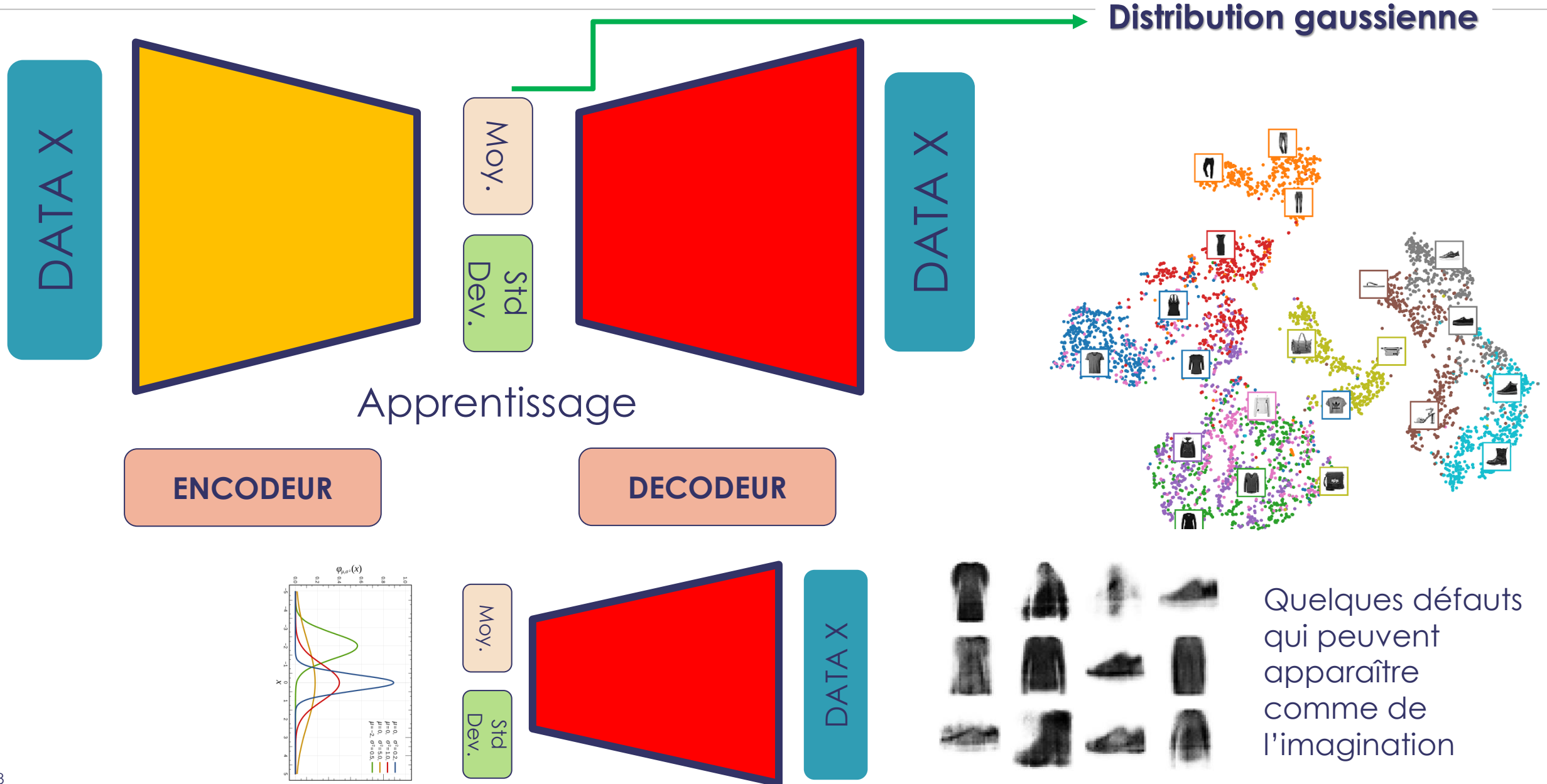


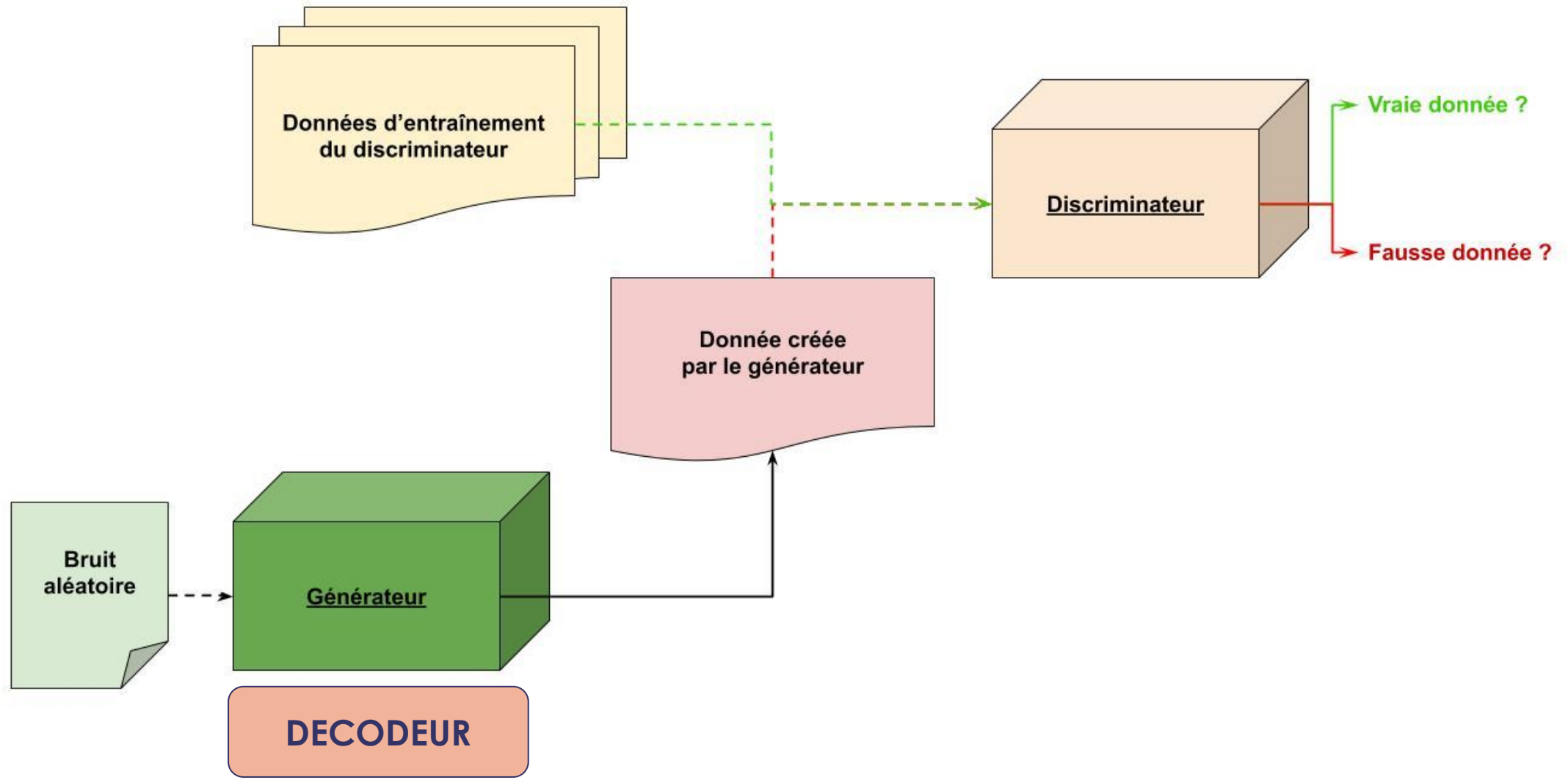
$$Y = X$$



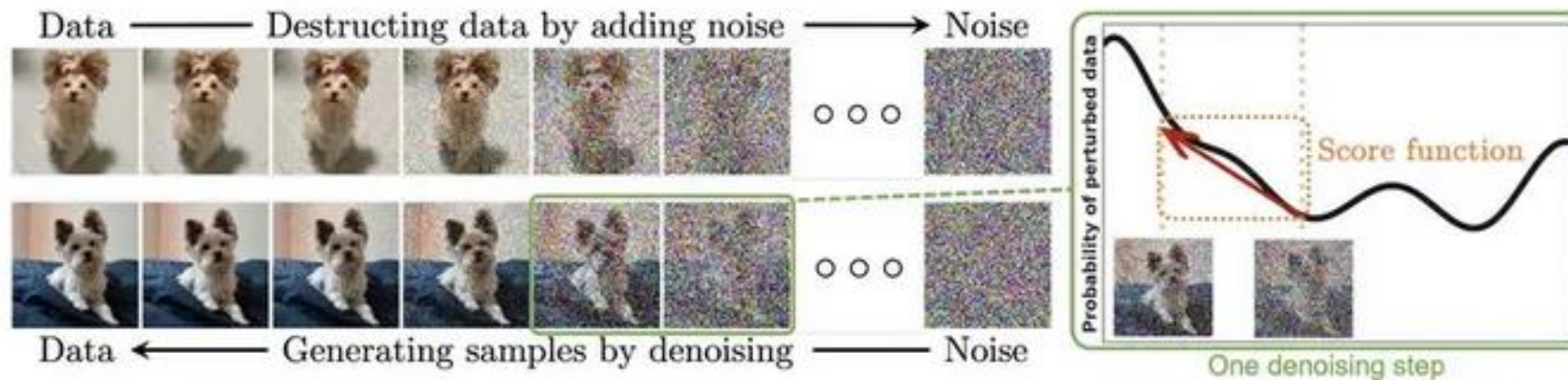
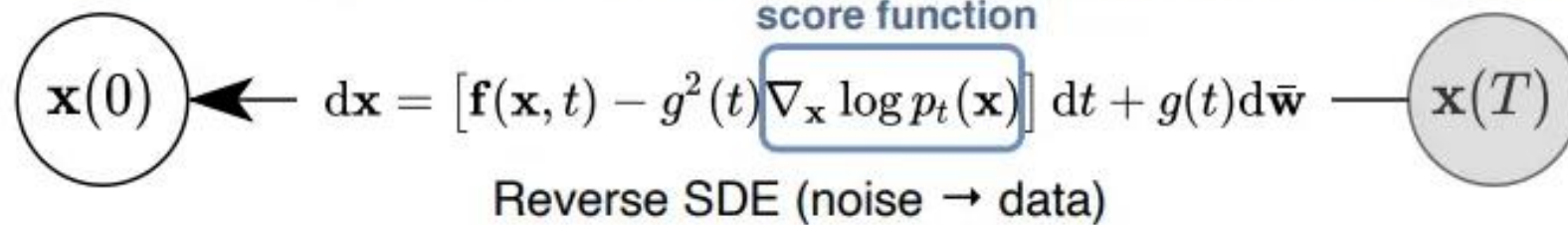
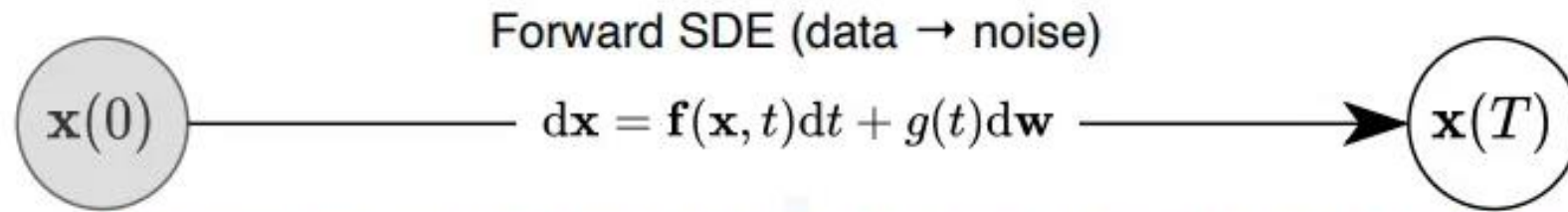
1. Reduction de dimensionalité
2. Extraction de caractéristiques
3. Débruitage d'images
4. Recherche d'images
5. Détection d'anomalies
6. Création de valeurs manquantes







➔ Amélioration du générateur de données grâce à un discriminateur de « fake »



- Texte vers Image



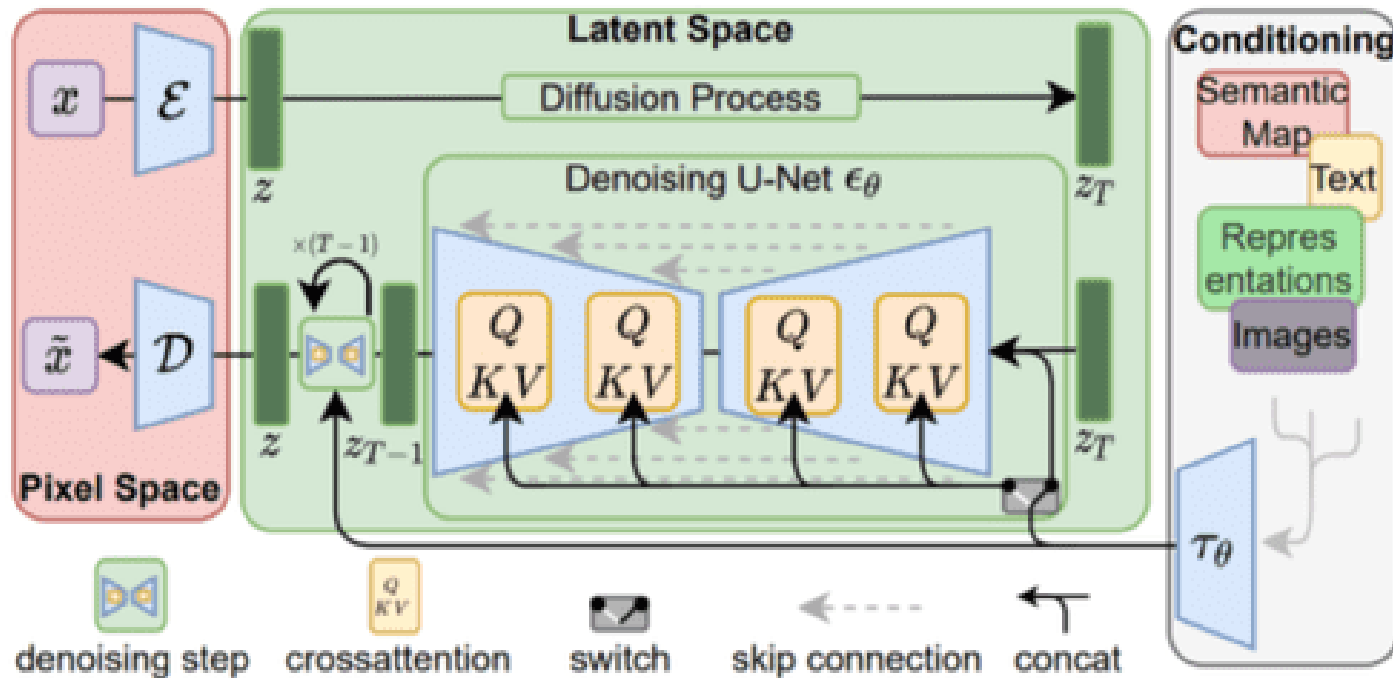
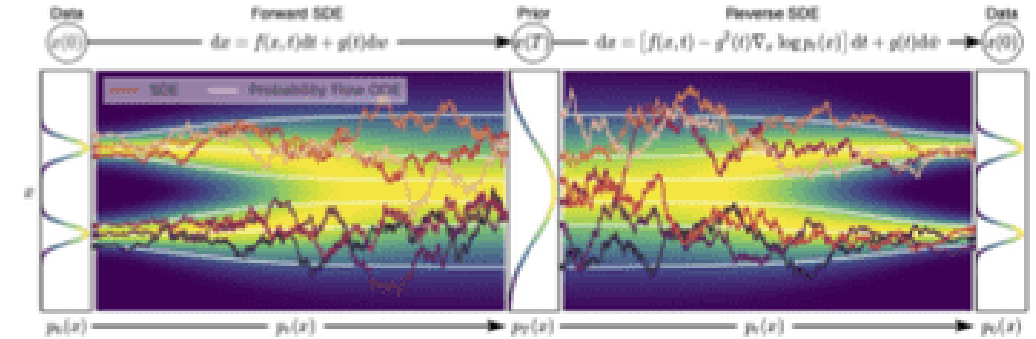
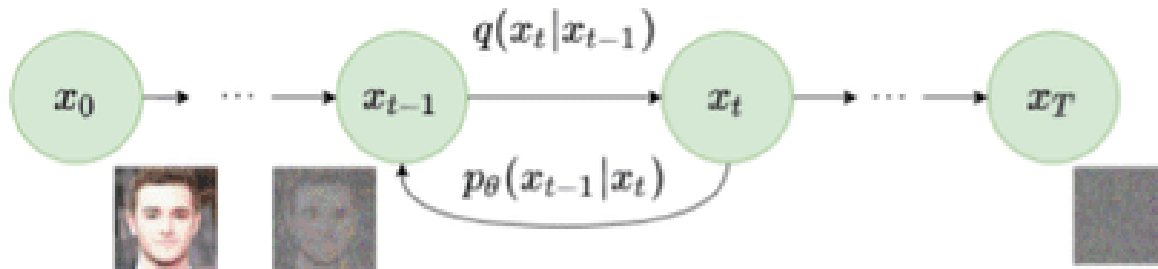
Source : <https://aithub.com/bentoml/stable-diffusion-bentoml/>



Source : DALL-E 2



Source : <https://ai.meta.com/blog/emu-text-to-video-generation-image-editing-research/>



Algorithm 1 Training

- 1: repeat
- 2: $x_0 \sim q(x_0)$
- 3: $t \sim \text{Uniform}(\{1, \dots, T\})$
- 4: $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
- 5: Take gradient descent step on $\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, t)\|^2$
- 6: until converged

Algorithm 2 Sampling

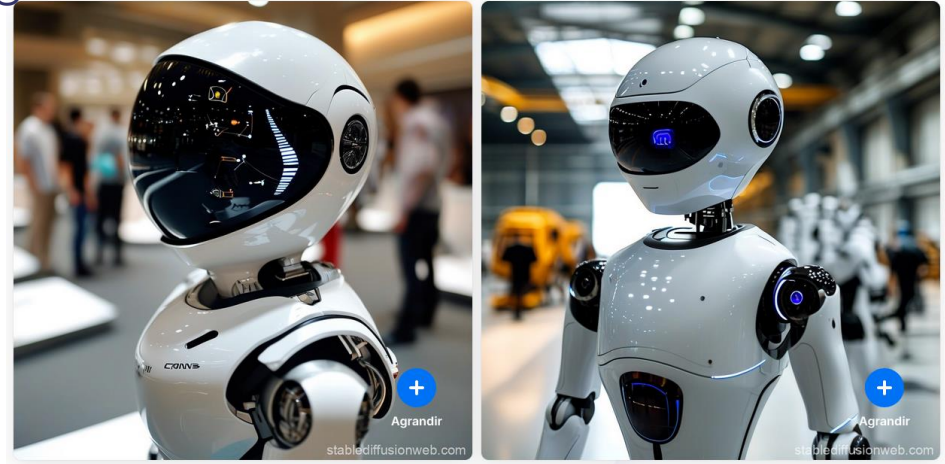
- 1: $x_T \sim \mathcal{N}(0, \mathbf{I})$
- 2: for $t = T, \dots, 1$ do
- 3: $z \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $z = 0$
- 4: $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z$
- 5: end for
- 6: return x_0

<https://openart.ai/create>



cours sur l'IA dispensés à des adultes dans le cadre professionnel avec des images de modèle d'IA CNN, RNN, DNN, GAN. L'image est réaliste.

<https://stablediffusionweb.com/fr/app/image-generator>



<https://stablediffusionweb.com/fr/app/image-generator>





- ➔ Recherchons quelle partie est importante
- ➔ Identifions les parties qui nécessitent notre attention



Question ? query



traitement du signal



Tous Images Vidéos Actualités Produits Livres Web Plus

Outils

Le traitement du signal **c'est la réalisation d'opérations sur le signal**. – Elaboration de signaux : Synthèse (de parole, de musique), modulation, codage. – Interprétation des signaux : filtrage, extraction/détection d'information, identification, analyse (spectrale ou temporelle) ou mesure.

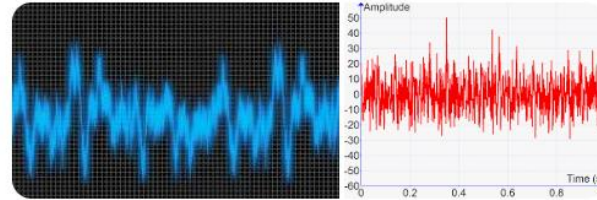


Le Mans Université

https://perso.univ-le Mans.fr/doc/cpda_signal PDF

Traitement du signal

Key2 ?



Key1 ?

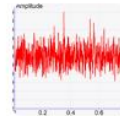


Wikipédia

https://fr.wikipedia.org/wiki/Traitement_du_signal

Traitement du signal

Le **traitement du signal** est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux.



LIAS Laboratory

<https://www.lias-lab.fr/fredericlaunay/Cours> PDF

Traitement du signal : les fondamentaux

Initialement destiné à extraire le signal dans un bruit lors de mesures (capteurs), le traitement du signal est largement appliqué en Télécommunication dans ...
50 pages

Key3 ?



FUN MOOC

<https://www.fun-mooc.fr/cours/introduction-au-traitement-du-signal>

Introduction au traitement du signal - Cours

Les thèmes abordés sont la **représentation temporelle et spectrale des signaux**, le filtrage, la numérisation et les applications.

Key4 ?

Traitement du signal

Discipline :

Le traitement du signal est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux. [Wikipédia](#)

Les internautes recherchent aussi



Apprentissage



Design



Communication



Stratégie de régulation

Commentaires

Réponses: Keys

Attention :
Meilleure Valeur calculée

Question ? query



traitement du signal



Tous Images Vidéos Actualités Produits Livres Web Plus

Outils

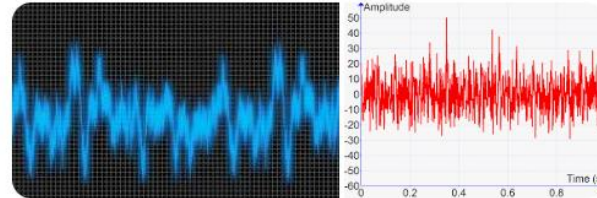
Le traitement du signal c'est la réalisation d'opérations sur le signal. – Elaboration de signaux : Synthèse (de parole, de musique), modulation, codage. – Interprétation des signaux : filtrage, extraction/détection d'information, identification, analyse (spectrale ou temporelle) ou mesure.



Le Mans Université

https://perso.univ-lemans.fr/doc/cpda_signal PDF

Traitement du signal



Key2 ? Valeur 2: Value

À propos des extraits optimisés Commentaires

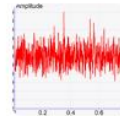


Wikipédia

https://fr.wikipedia.org/wiki/Traitement_du_signal

Traitement du signal

Le **traitement du signal** est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux.



LIAS Laboratory

<https://www.lias-lab.fr/fredericlaunay/Cours> PDF

Traitement du signal : les fondamentaux

Initialement destiné à extraire le signal dans un bruit lors de mesures (capteurs), le traitement du signal est largement appliqué en Télécommunication dans ... 50 pages

Key3 ?

Valeur 3: Value



FUN MOOC

<https://www.fun-mooc.fr/cours/introduction-au-traitement-du-signal>

Introduction au traitement du signal - Cours

Les thèmes abordés sont la représentation temporelle et spectrale des signaux, le filtrage, la numérisation et les applications.

Key4 ?

Valeur 4: Value

Traitement du signal

Discipline :

Le traitement du signal est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux. Wikipédia >

Les internautes recherchent aussi



Apprentissag
e...



Design



Communicat
ion



Stratégie de
régulation

Commentaires

Réponses: Keys

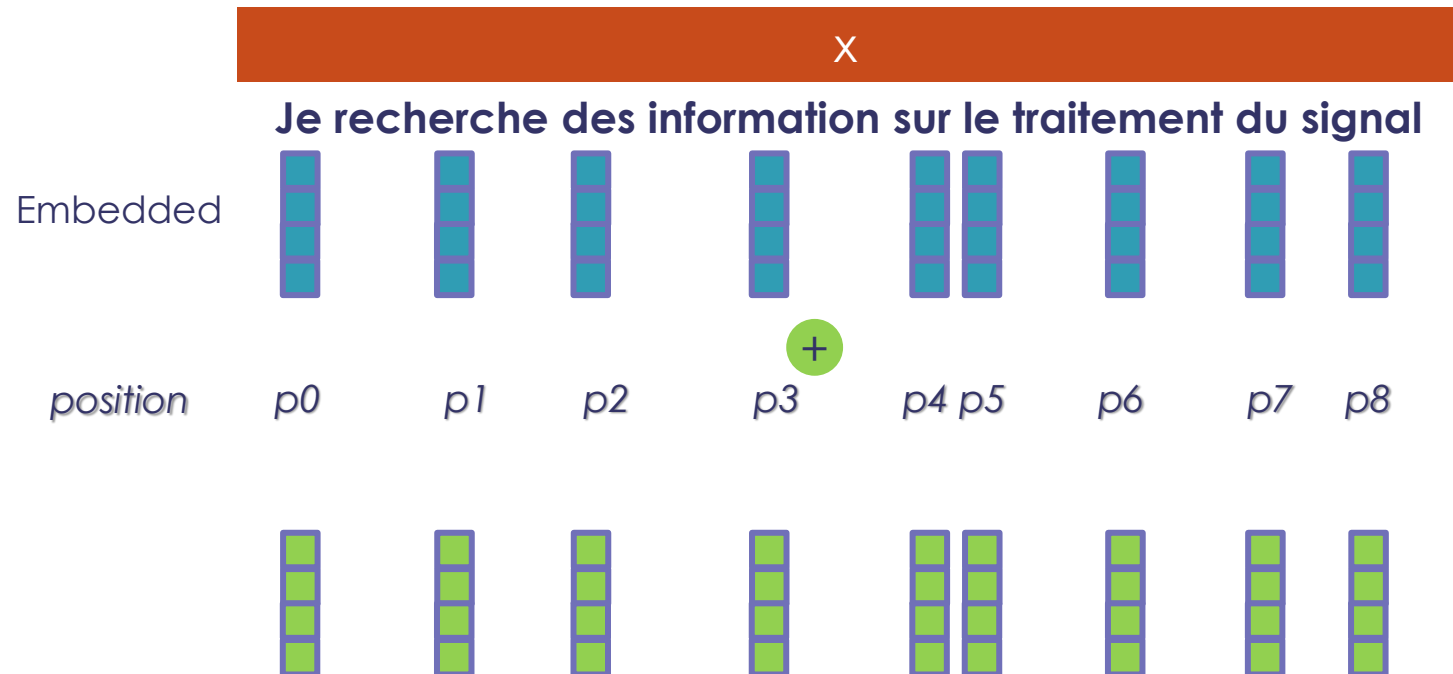
Key1 ?

Valeur 1: Value

Attention :
Meilleure Valeur calculée

L'attention est tout ce que nous avons besoin

1. Encoder la position et la sémantique des informations → Embedded
2. Extraire par calcul le trinôme (query, key, value) pour la recherche
3. Calculer les poids de l'attention
4. Extraire les caractéristiques avec une grande attention



→ Encodage sémantique + position

1. Encoder la position et la sémantique des informations → Embedded

2. Extraire par calcul le trinôme (query, key, value) pour la recherche

3. Calculer les poids de l'attention

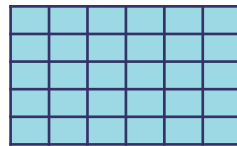
4. Extraire les caractéristiques avec une grande attention

Position

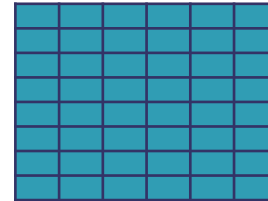


X

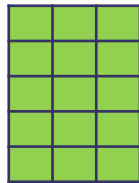
NN



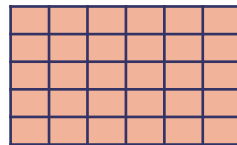
=



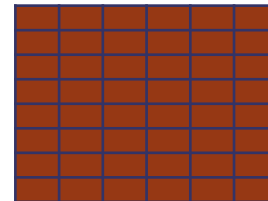
Q: Query



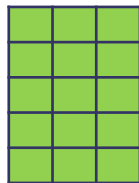
X



=



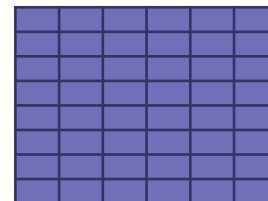
K: Key



X



=



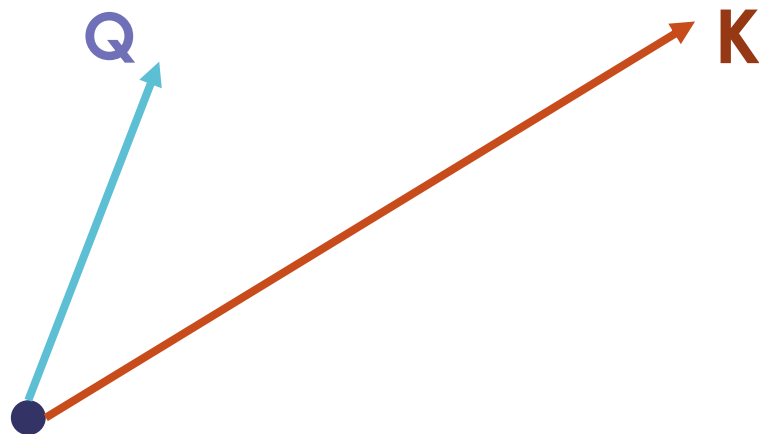
V: Value

1. Encoder la position et la sémantique des informations → Embedded
2. Extraire par calcul le trinôme (query, key, value) pour la recherche
3. Calculer les poids de l'attention

Cosine similarity = métrique de similarité

$$\frac{Q \cdot K^T}{MaE}$$

4. Extraire les caractéristiques avec une grande attention



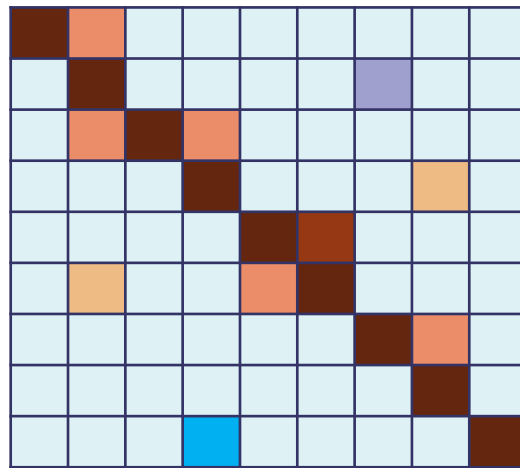
Je
recherche
des
information
sur
le
traitement
du
signal

	Je	recherche	des	information	sur	le	traitement	du	signal
Je	dark	orange							
recherche		dark					purple		
des		orange	dark	orange					
information				dark				orange	
sur					dark	dark			
le		orange			orange	dark			
traitement							dark	orange	
du								dark	
signal				blue					dark

$$\text{softmax}\left(\frac{Q \cdot K^T}{MaE}\right)$$

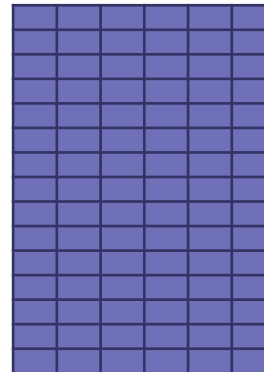
Poids d'attention

1. Encoder la position et la sémantique des informations → Embedded
2. Extraire par calcul le trinôme (query, key, value) pour la recherche
3. Calculer les poids de l'attention
4. Extraire les caractéristiques avec une grande attention



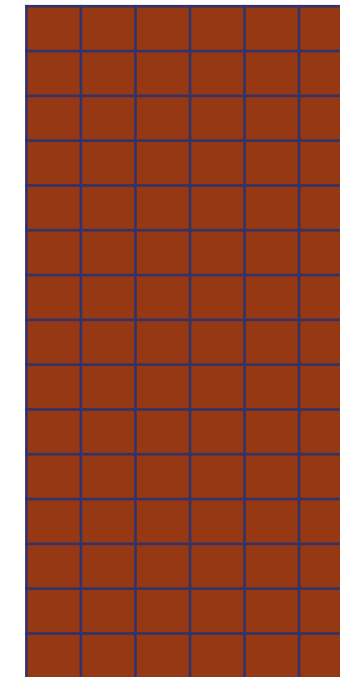
Poids d'attention

x



Value

=



Sortie

$$\text{softmax}\left(\frac{Q \cdot K^T}{MaE}\right) \cdot V$$

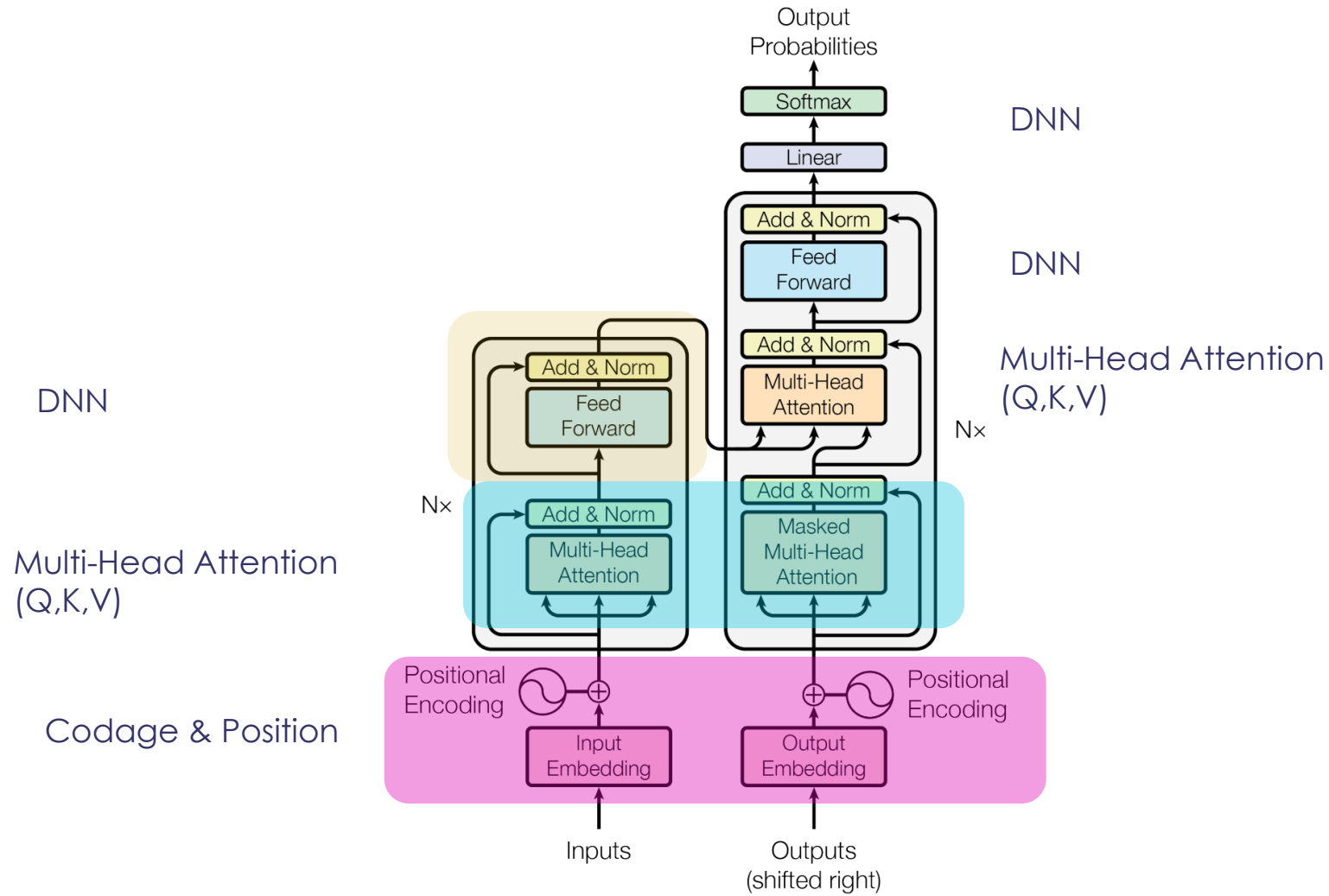
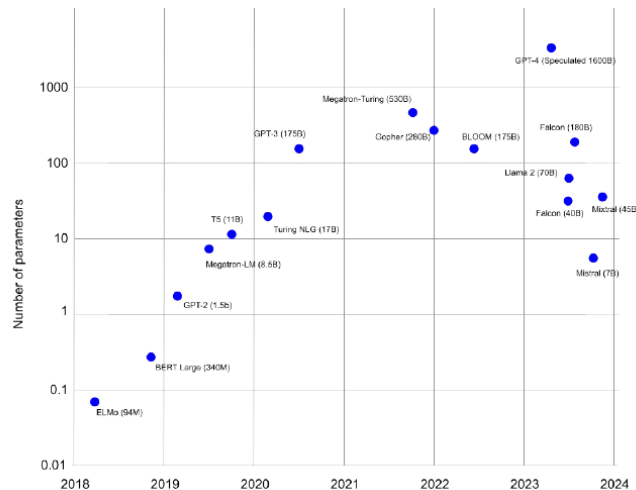
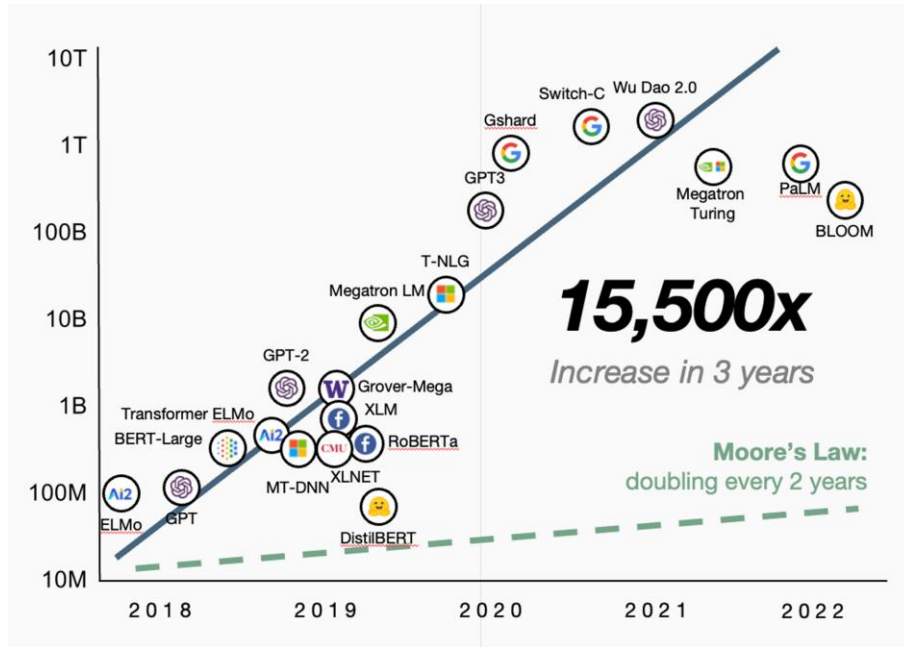
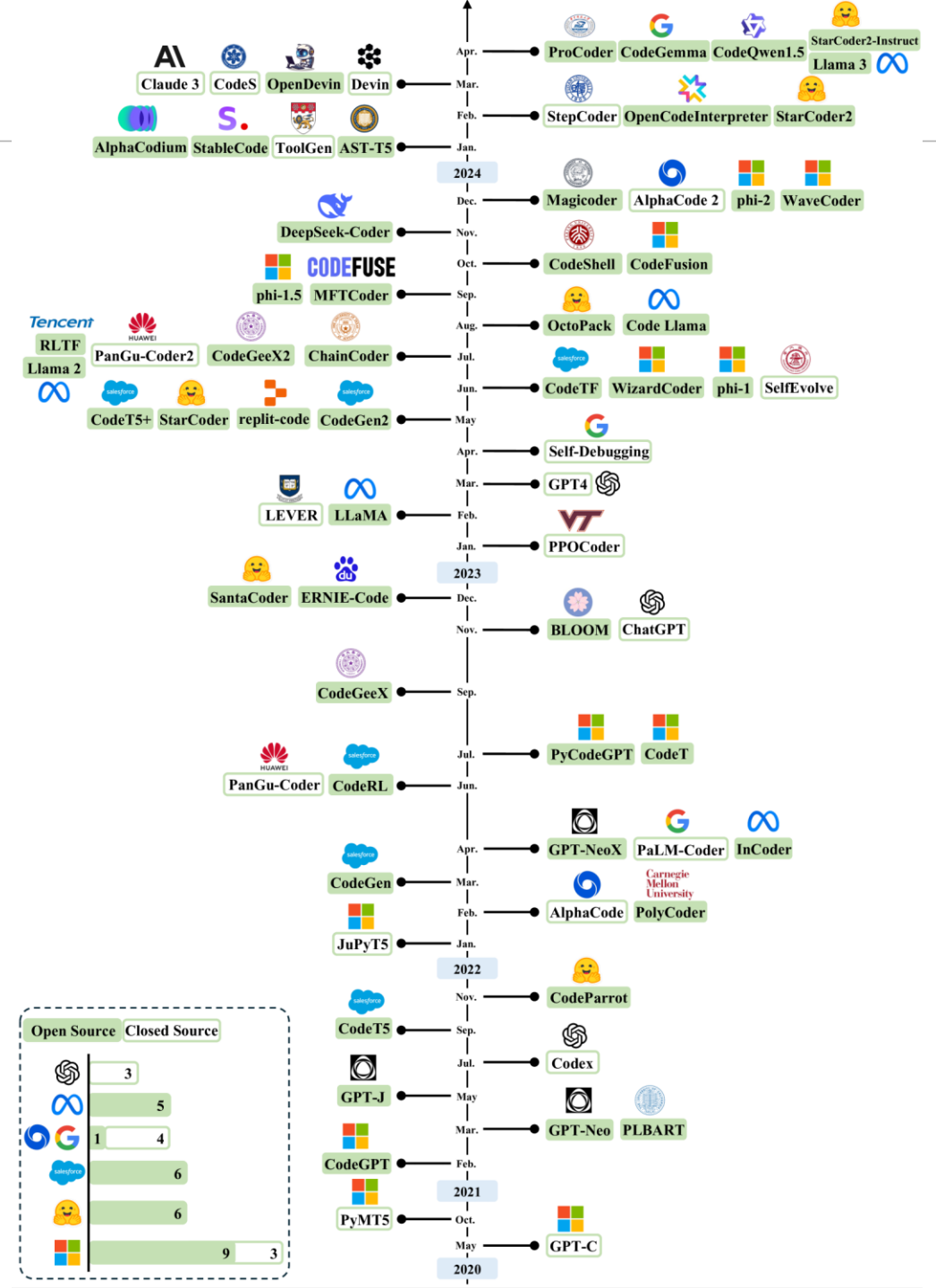


Figure 1: The Transformer - model architecture.

Les transformers !

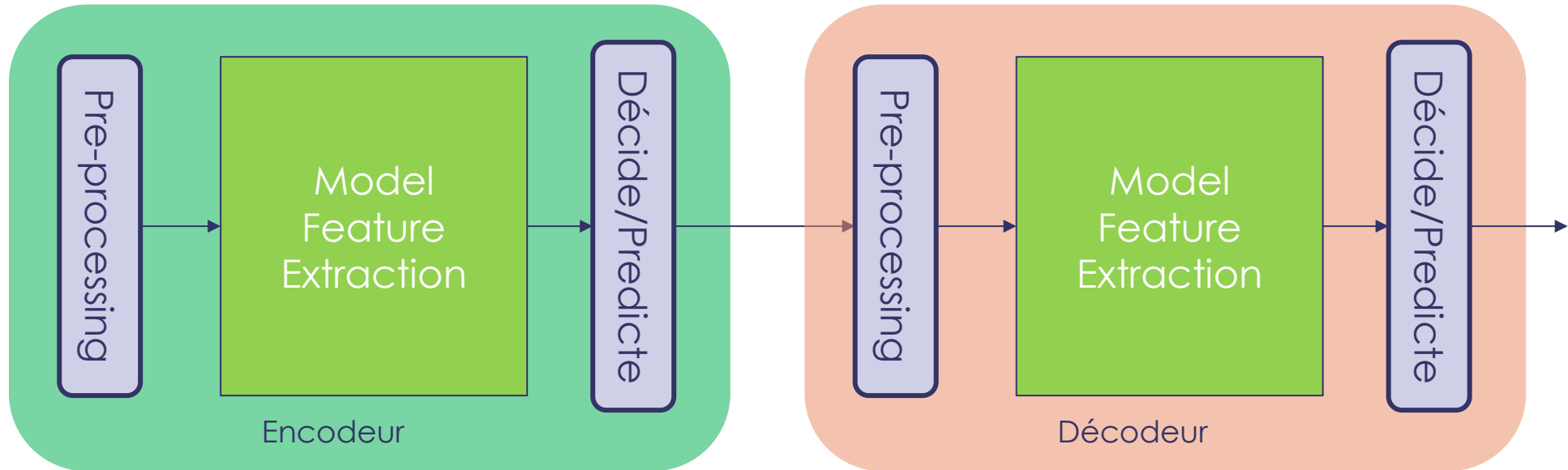


➔ Vers l'IAG ?





BERT, ChatGPT



Le modèle d'attention (autoencoders avec contraintes) → très demandant mais des progrès

Modèle conversationnel :

ChatGPT

Mistral

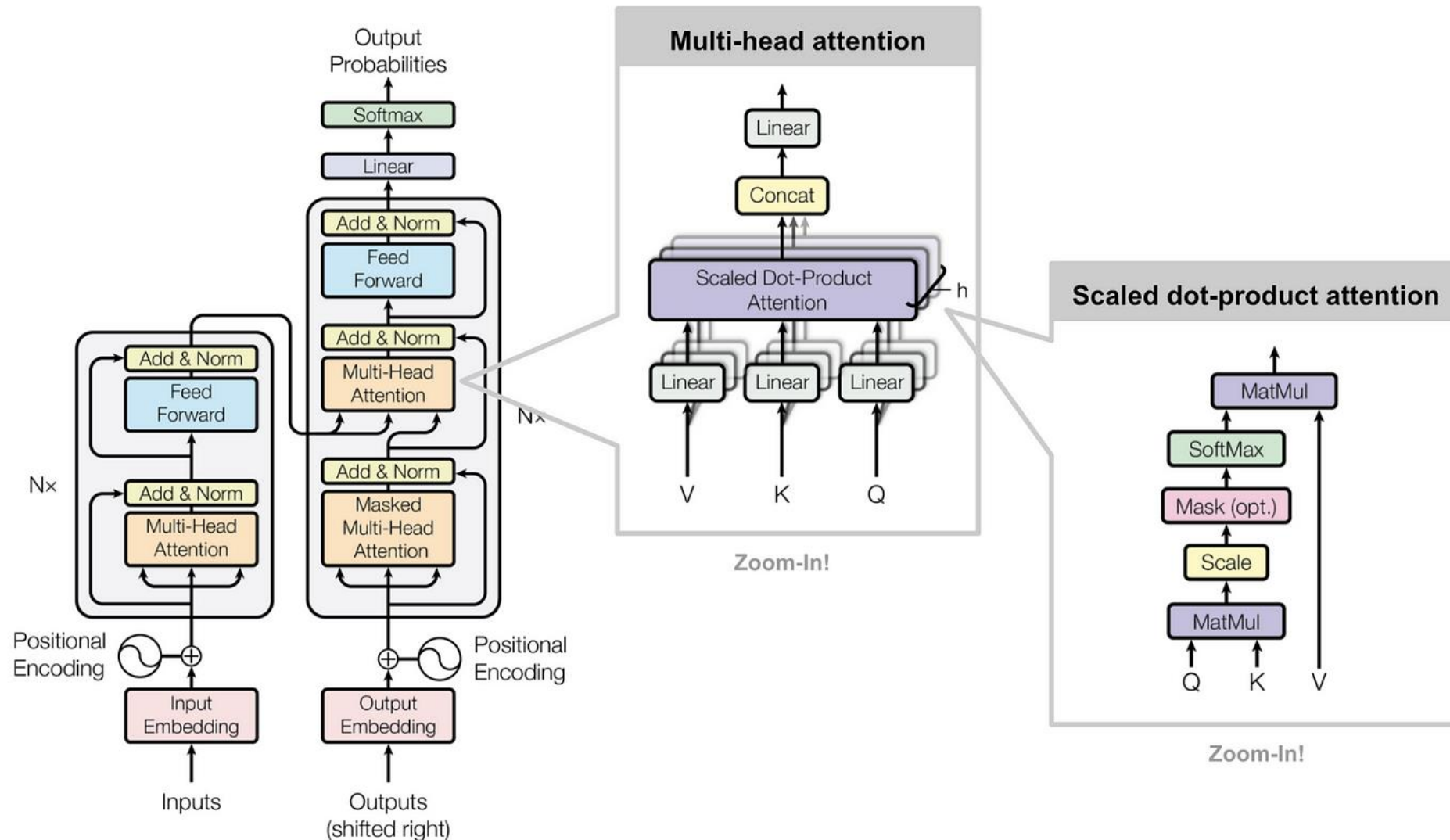
Bloom

LLama2

LLM sont très puissants

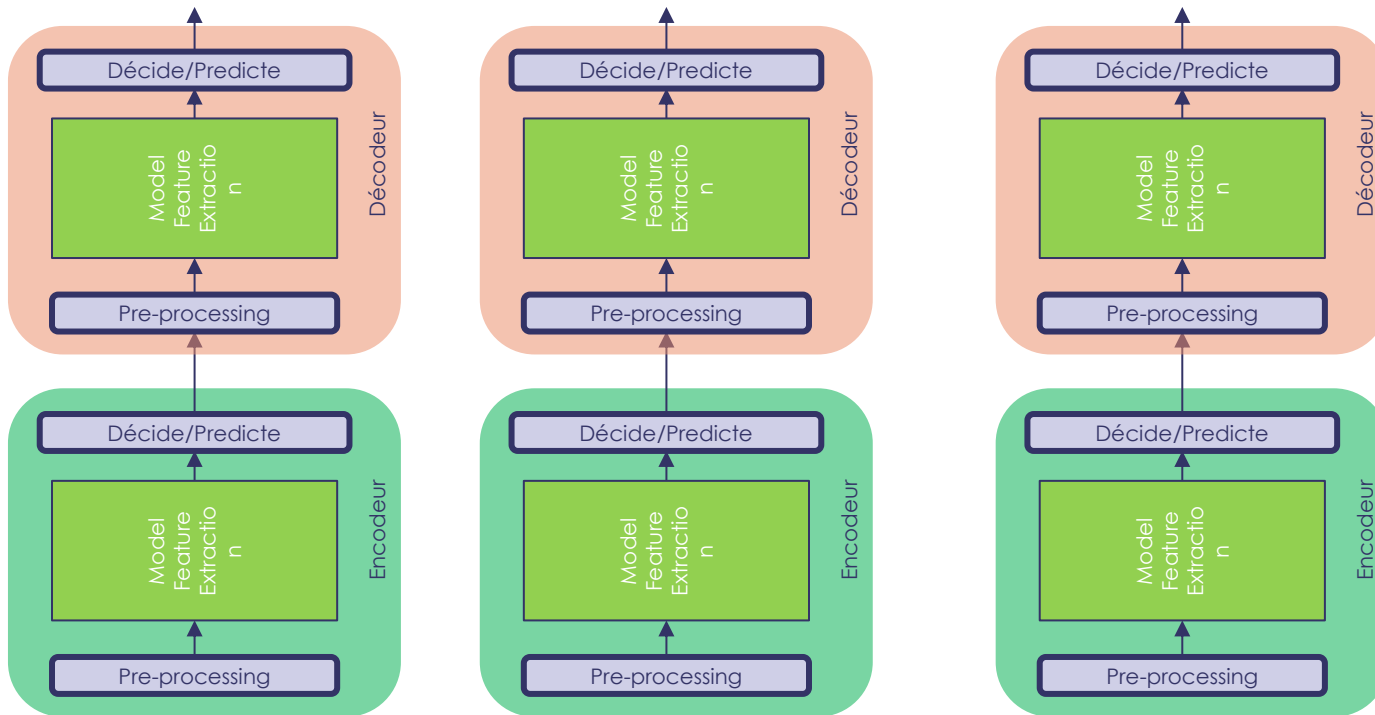
Mais, pas d'inference possible en IoT (trop de puissance de calcul et de mémoire)

Utilisation des LLMs pour améliorer la **labellisation des données**, augmenter la **quantité de données**.



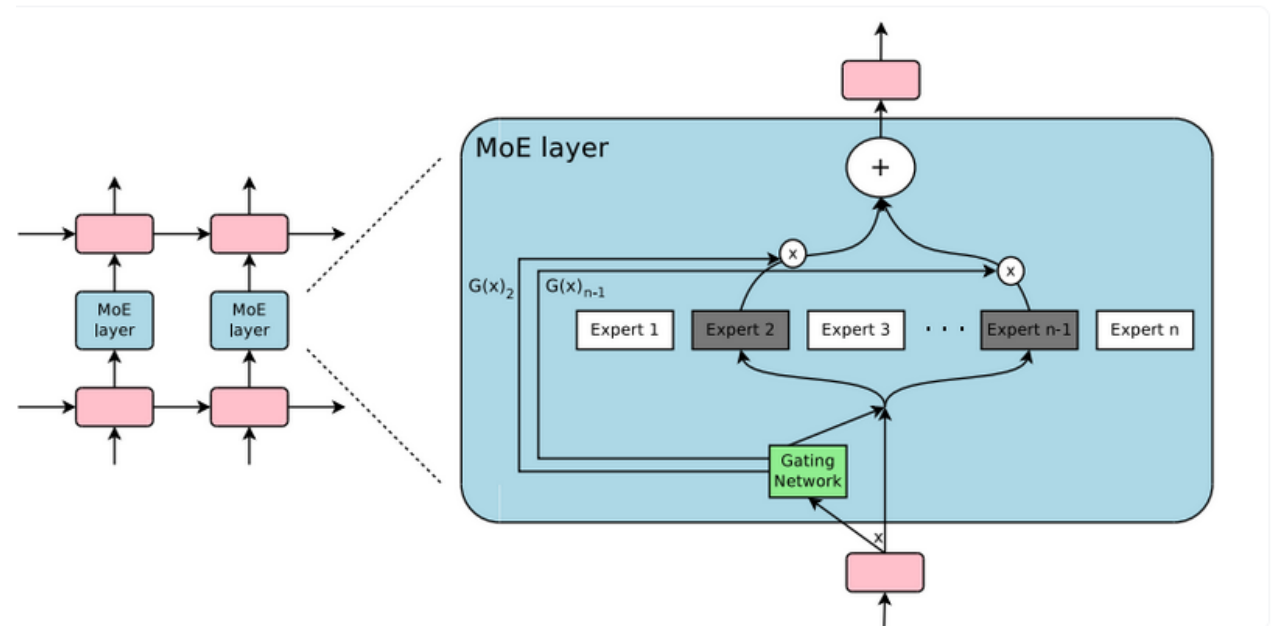
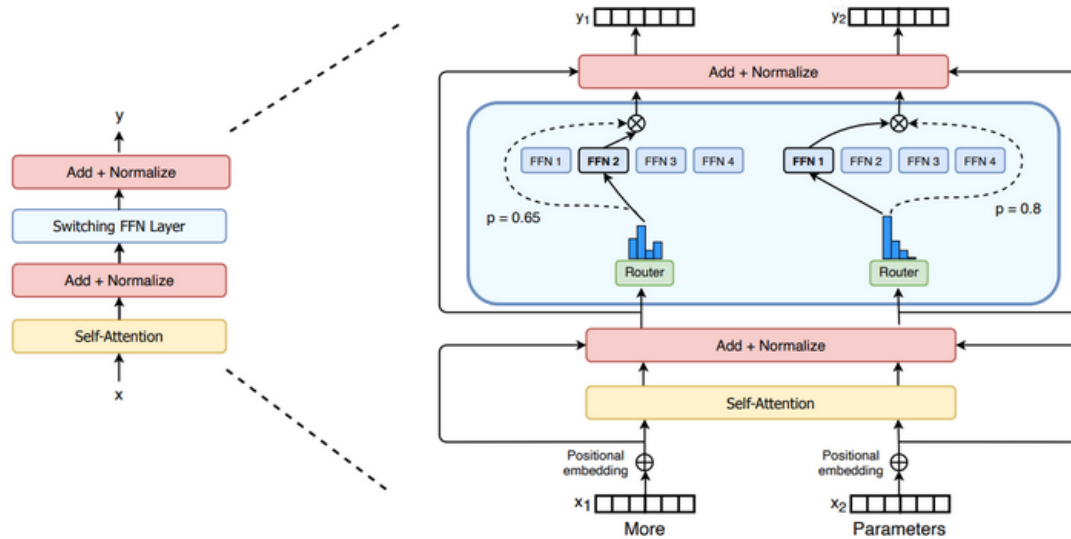


Transformer spécialisés



Trop difficile à
entraîner

Mixture of Experts => Comment entrainer plus efficacement un large modèle



MoE layer from the Outrageously Large Neural Network paper

What do you need to train a large language model ?

A truckload of data

Transformers are ravenous. You need to feed them with a substantial and hard-to-come-by dataset.



A mighty compute infrastructure

GPUs with high throughput and large VRAM can execute this training in a reasonable amount of time.



A copious amount of electricity

Storage, memory and compute power consume a lot of electricity.



An abundance of manpower

Cleaning the dataset, making experiments, monitoring SOTA advancements is a lot of work.



Single Agent Interaction



Multi Agent Interaction

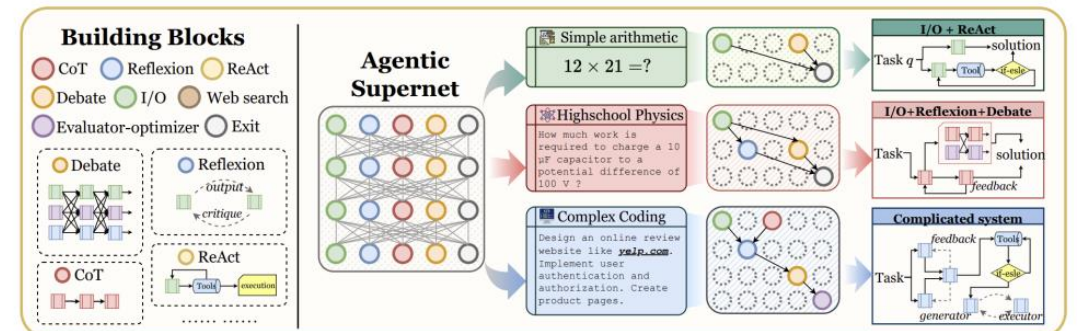
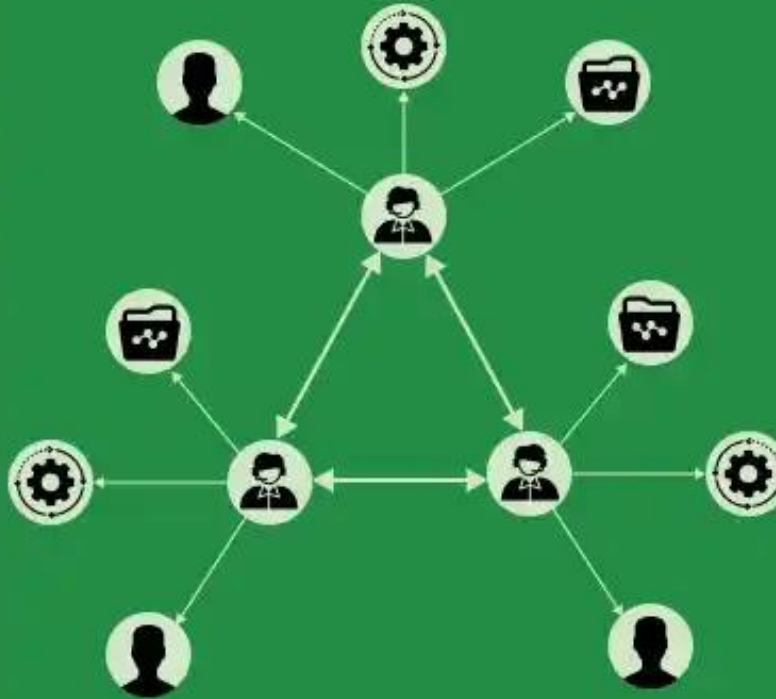
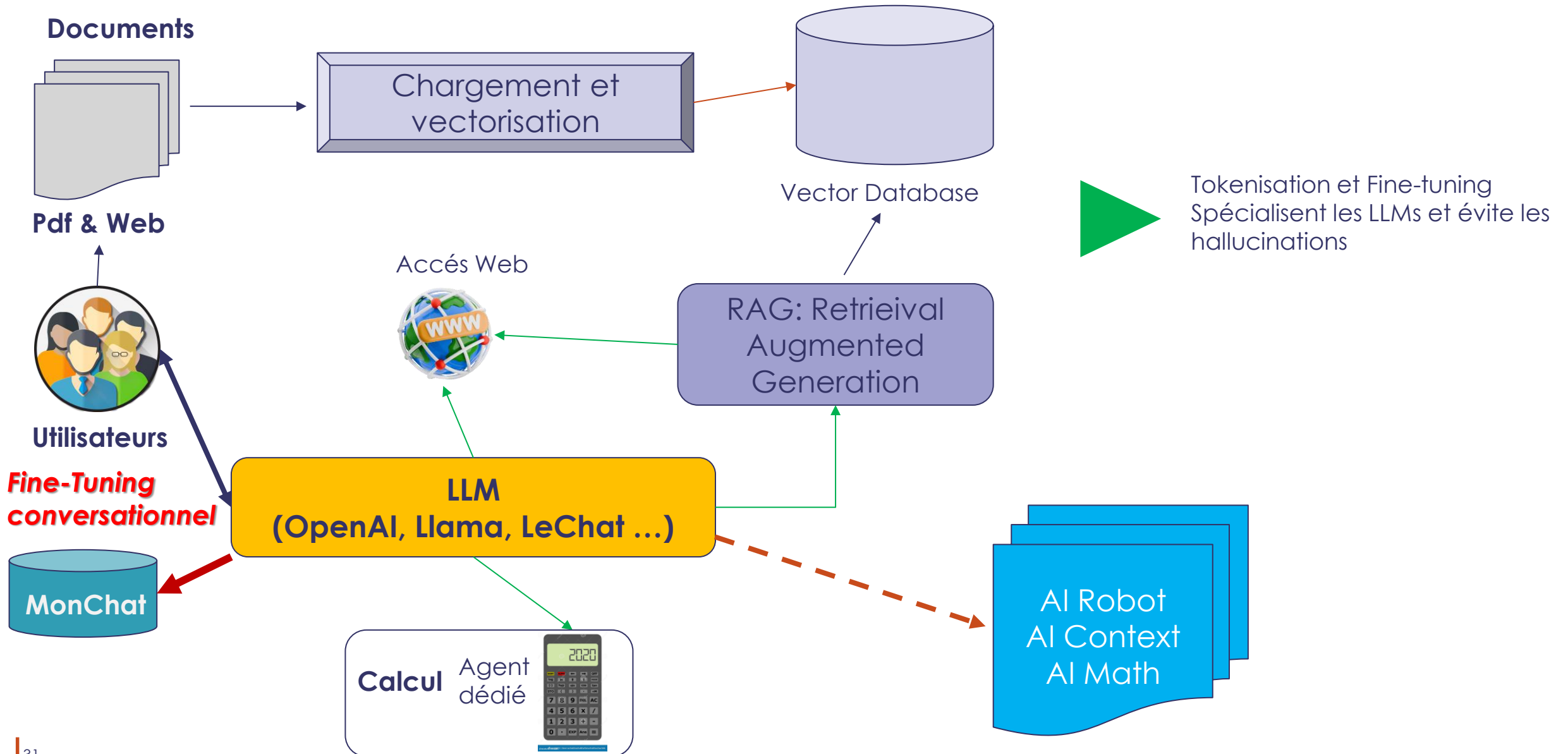
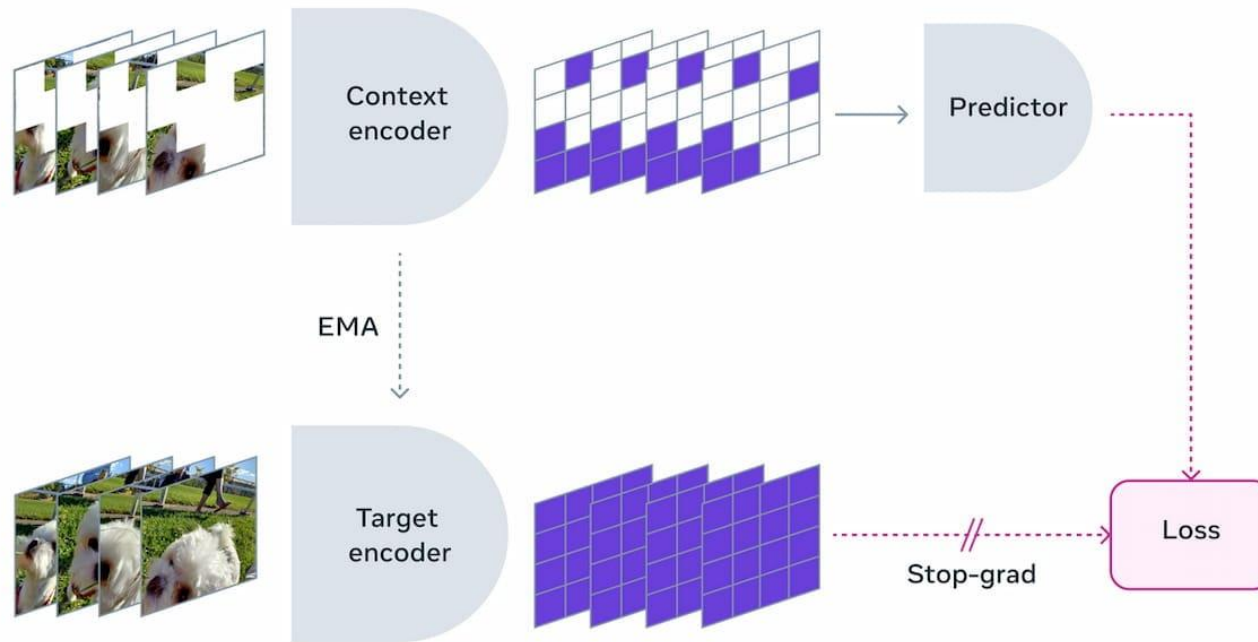


Figure 1. (Left) The building blocks of MaAS; (Right) When confronting different queries, the agentic supernet adaptively samples tailored multi-agent architecture in a query-dependent manner.



Apprendre avec les vidéo comme un humain !



- 5 à 10 ans de dev!
- StartUp AMI (LeCun)
- 1 Milliards de \$ pour une idée
- Ca marche pour des robots

$9 \cdot 10^{14}$ Octets Web : Apprentissage par du Texte

- 10^{14} Octets vu par un enfant de 4 ans
- Quelques semaines pour apprendre à conduire
- Créer un modèle du monde avec des videos pour apprendre le contexte/les concepts et non pixels par pixels

Conclusion

- Les réseaux de neurones
- Les bases DNN, CNN et RNN
- Les modèles génératifs
- La phase d'apprentissage
- Gestion des informations sous forme numériques

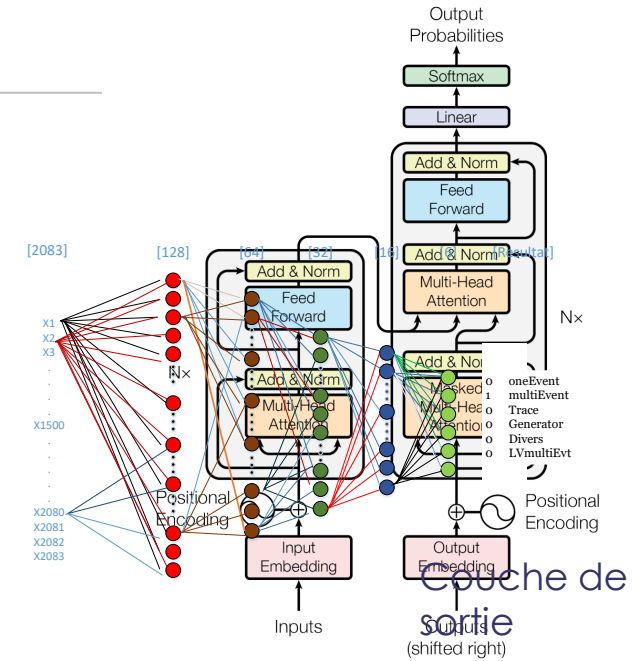
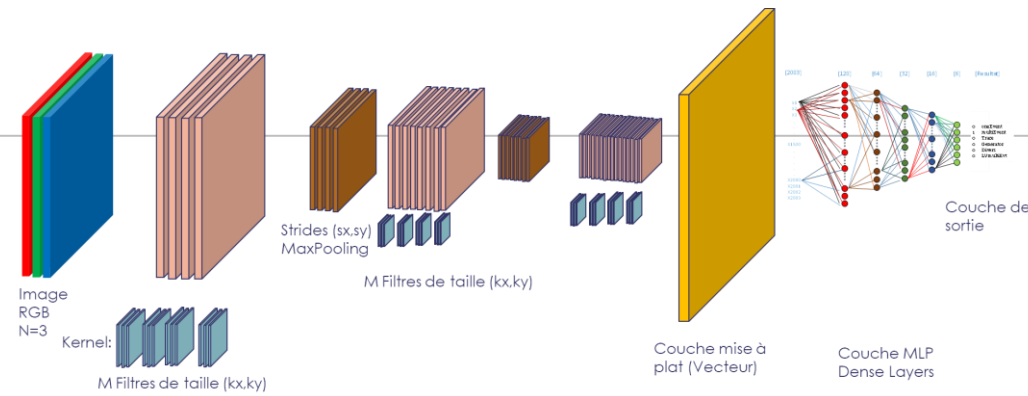
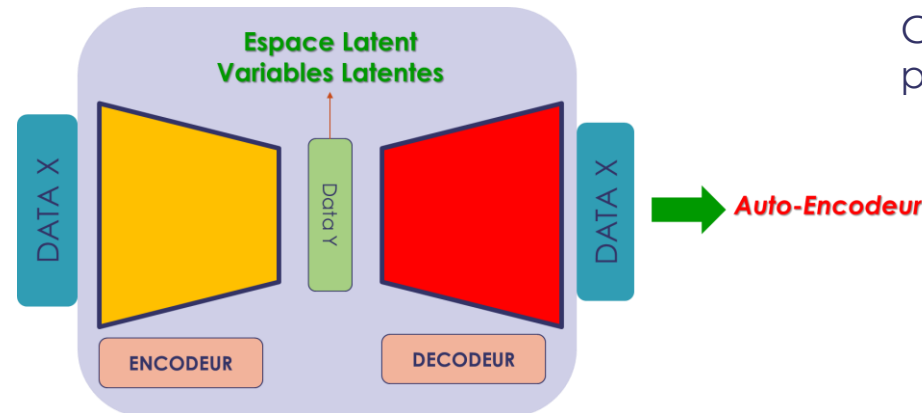
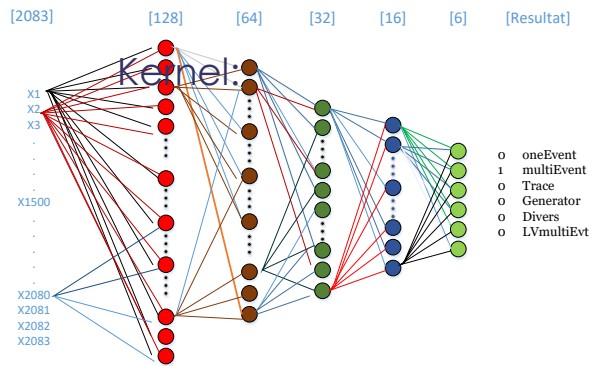


Figure 1: The Transformer - model architecture.



Couche mise à plat (Vecteur)

Couche MLP Dense Layers

