IA générative démystifiée

Laure Soulier

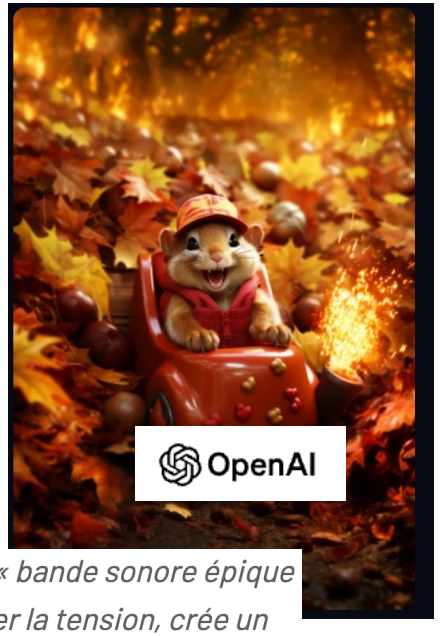
Journée Capsule — 20 juin 2024

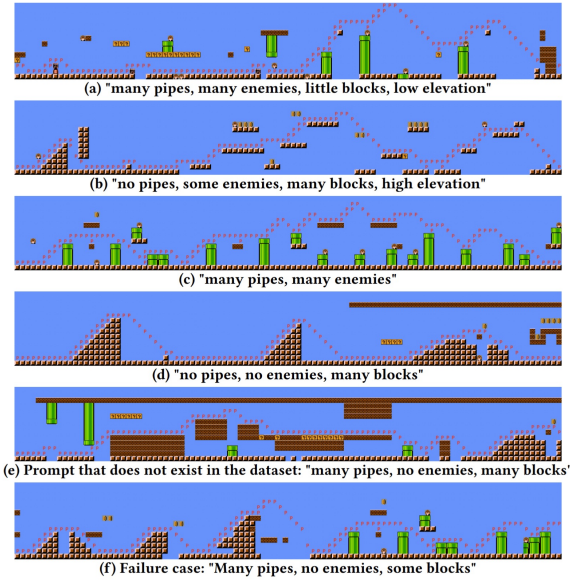
Le pouvoir de l'IA générative

LA Ecris un texte de 5 phrases sur Jian-Yun Nie

OpenAI

Jian-Yun Nie est un chercheur et professeur émérite en informatique. Il est connu pour ses contributions significatives dans le domaine du traitement de l'information et de la recherche d'information. Ses travaux portent notamment sur l'exploration de textes, la recherche d'informations multilingues et la fouille de données. En tant que chercheur de renommée mondiale, Jian-Yun Nie a publié de nombreux articles scientifiques et a participé à des conférences internationales. Ses recherches ont eu un impact important dans le domaine de l'informatique et continuent d'influencer de nouvelles avancées technologiques.





Pour en avoir un aperçu, voici l'extrait généré pour le prompt *« bande sonore épique utilisant des instruments orchestraux. Le morceau fait monter la tension, crée un sentiment d'urgence. Un chœur a cappella chante à l'unisson, ce qui crée un sentiment de puissance et de force »* :

00:04 00:30

https://www.blogdumoderateur.com/musiclm-nouvel-outil-ia-google-generer-musique/
And (Sudhakaran et al 2023)

IA, qu’èsaquo?

Intelligence(s) Artificielle(s)

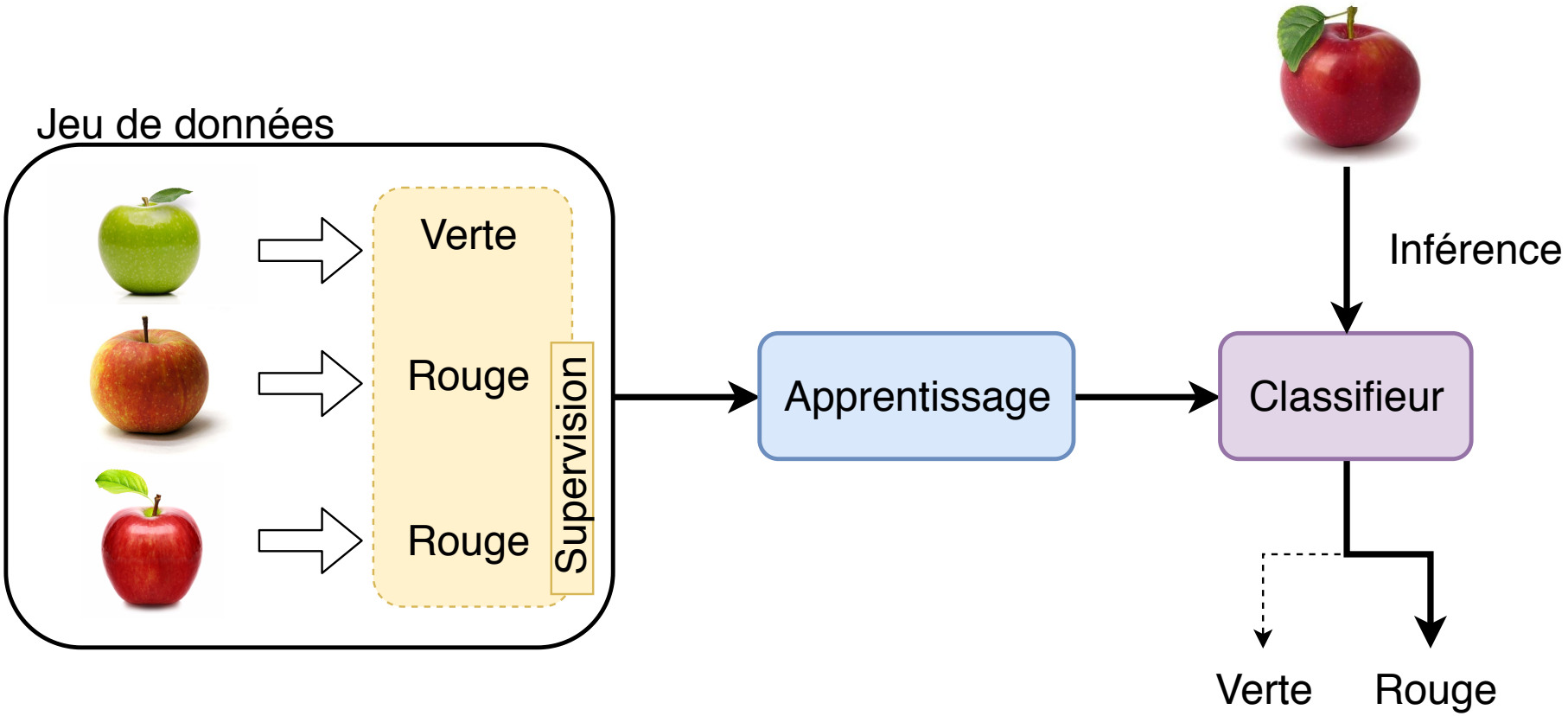
IA : programmes informatiques qui s'adonnent à des **tâches** qui sont, pour l'instant, accomplies de façon plus satisfaisante par des **êtres humains** car elles demandent des **processus mentaux de haut niveau**.
Marvin Lee Minsky

Input (A)	Output (B)	Application
email	spam? (0/1)	spam filtering
audio	text transcript	speech recognition
English	Chinese	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info	position of other cars	self-driving car
image of phone	defect? (0/1)	visual inspection

Ne pas confondre la **NAI (Narrow Artificial Intelligence)**, dédiée à une tâche et la **GAI (General AI)** qui remplace l'humain dans des systèmes complexes.
Andrew Ng

L'intelligence, aujourd'hui, c'est beaucoup…

→ L'apprentissage par l'exemple (= machine learning)



Machine learning
Deep learning

Réseaux de neurones
SVM

Ingrédients de l'IA

Données Capteurs

Cadre logiciel

Stockage & Calcul

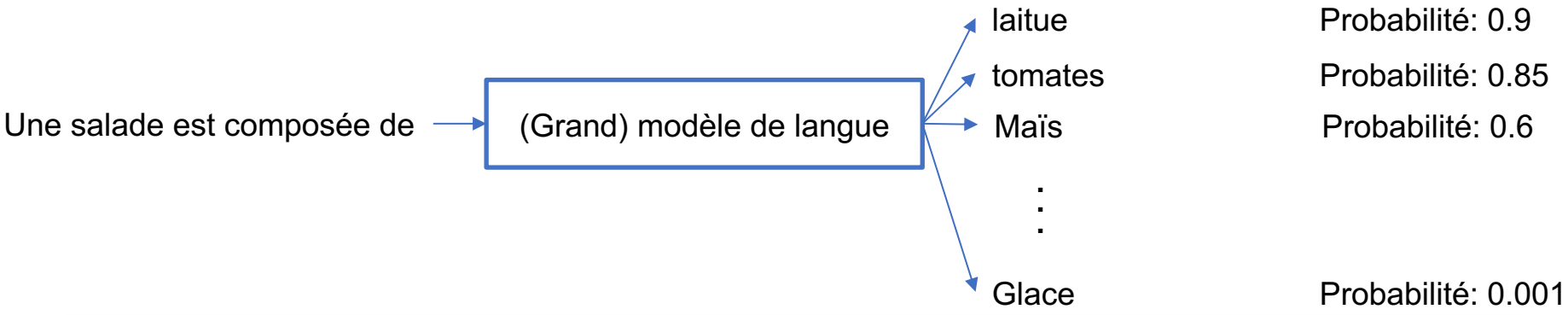
De la détection d’une pomme à ChatGPT

(Grands) Modèles de Langue

SCIENCES SORBONNE UNIVERSITÉ

Principe :

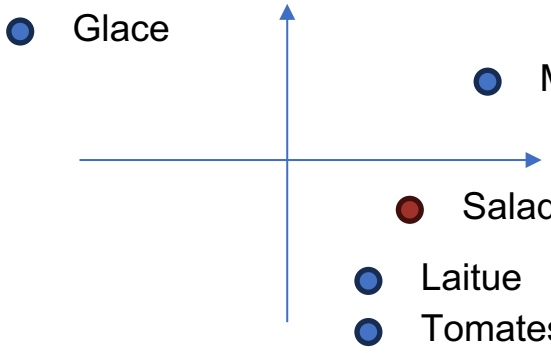
- Modeliser la probabilité d'une séquence $x_1, x_2, \ldots, x\_n$
- Les items peuvent être des mots, des caractères, des ngrams/bouts de mots, etc



Etant donné une séquence $x_1, x_2, \ldots, x_{n-1}$, quelle est la probabilité du prochain item x_n?

$$P(x_n | x_1, x_2, \ldots, x_{n-1})$$

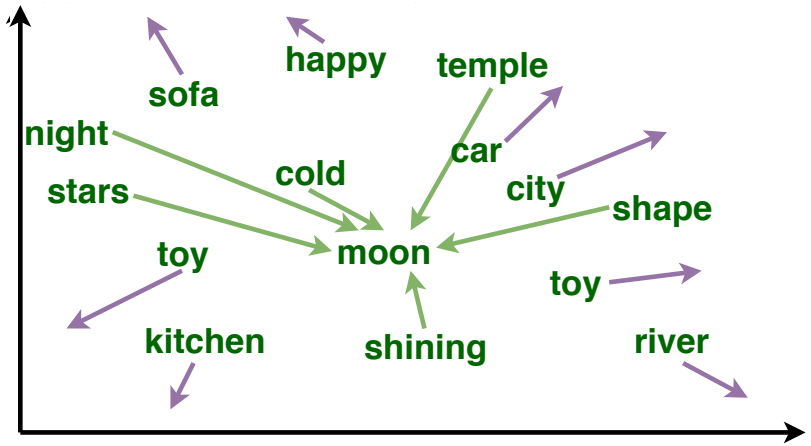
Sémantique, representation, espace latent



Salade = (0.3, 0.2, 0.45, -0.1, -0.3)
Laitue = (0.2, 0.1, 0.38, -0.5, -0.4)
…
Glace = (-0.9, -0.3, -0.5, 0.8, 0.7)

Apprentissage de représentations

→ Algorithme Word2Vec



he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
they love the sun , the moon and the stars . None of
the light of an enormous moon . The plash of flowing w
man 's first step on the moon ; various exhibits , aer
the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind

2000 — Modèle pionnier de Bengio

2012 — Word2Vec, FastText, …

2014 — Doc2Vec, FastSent, …

2017 — Représentations contextuelles Transformer networks Bert, T5, GPT, …

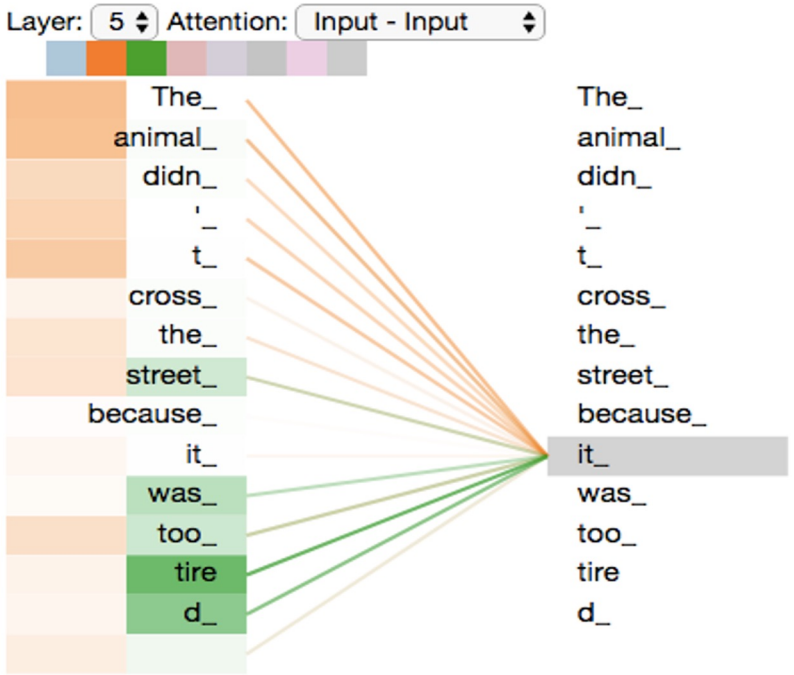
(Large) Language Models

Transformer (2017)

Un encoder-decoder avec :

- Environ 65 millions de paramètres (maintenant plus)
- Plusieurs blocs successifs
- Des têtes parallèles

… qui estime des representations contextuelles des items avec l'attention propre



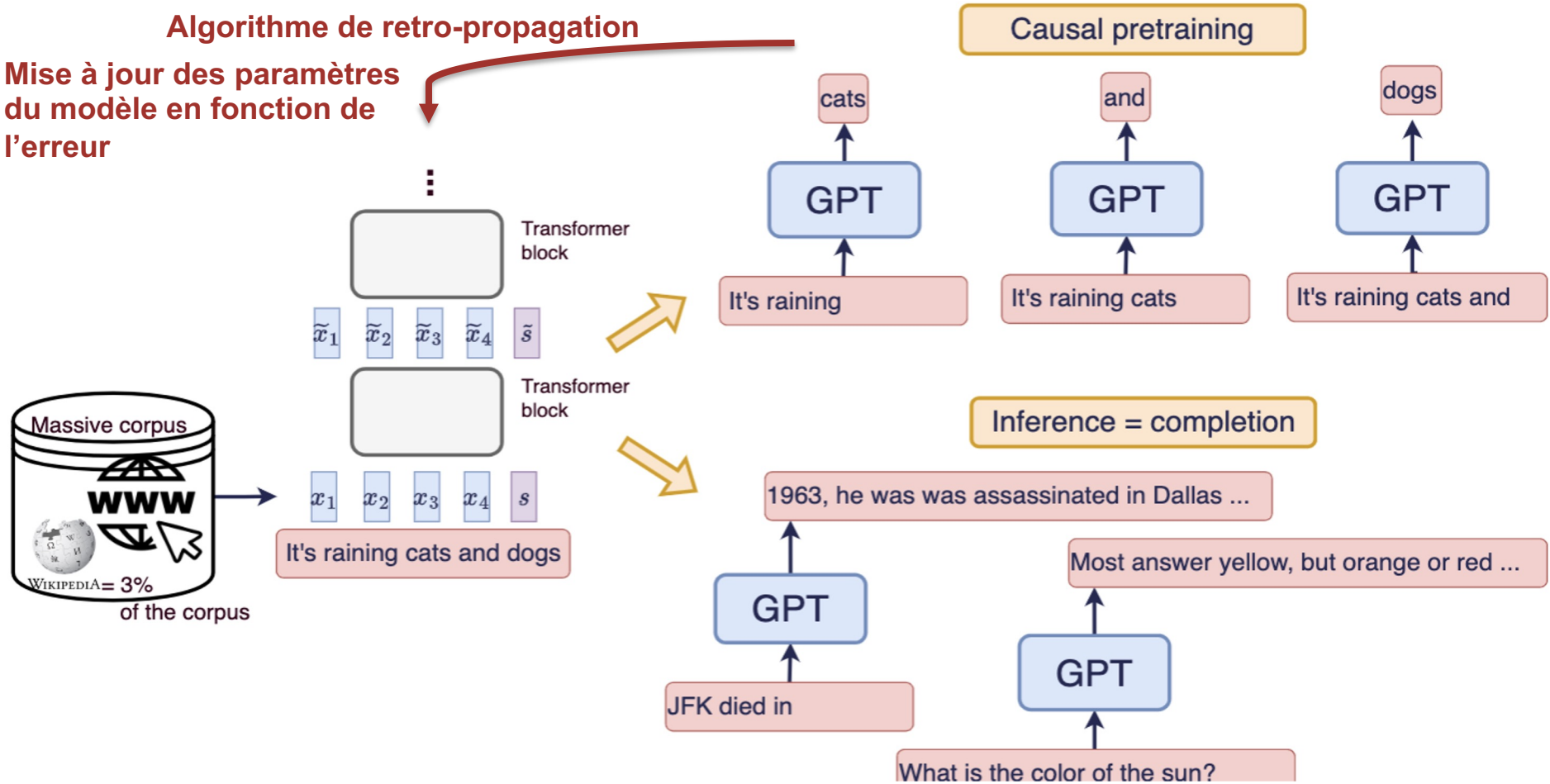
Distinguer Washington/city de Washington/man

(Vaswini et al 2017)

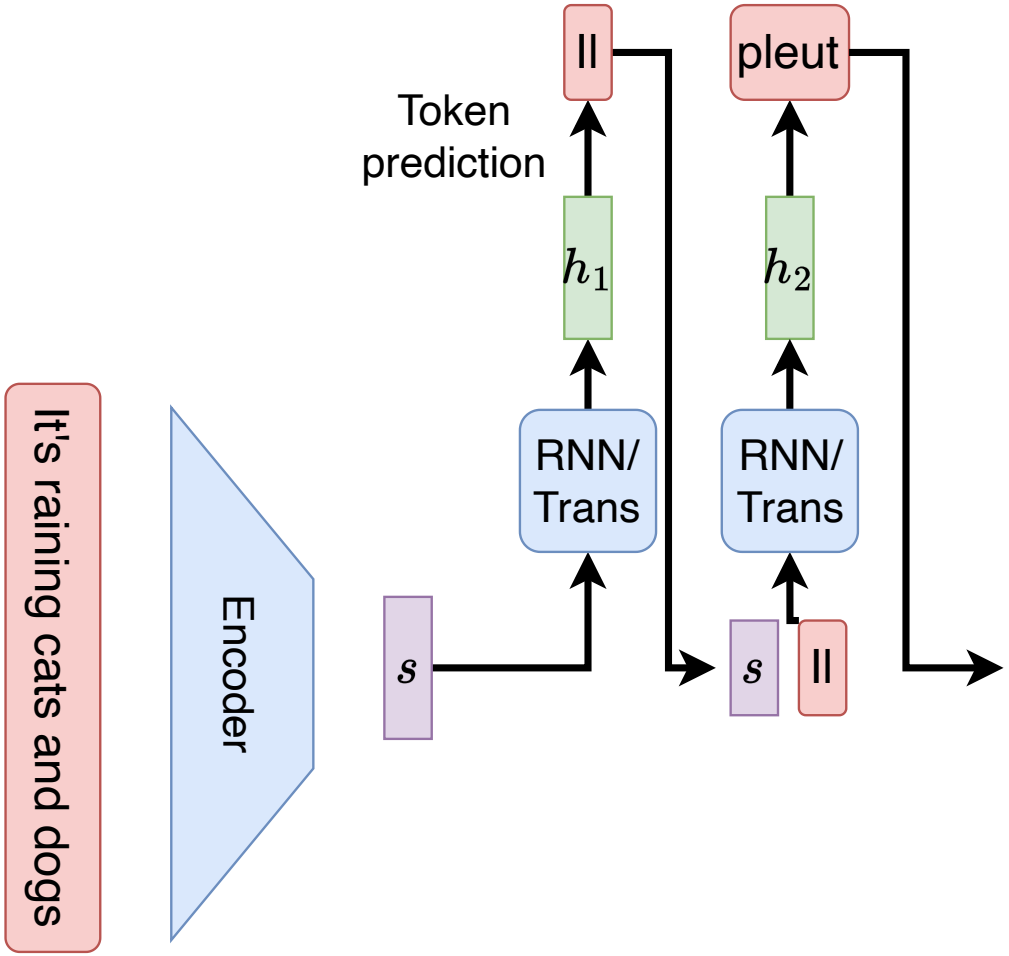
Entraîner et utiliser les gros modèles de langue

SCIENCES SORBONNE UNIVERSITÉ

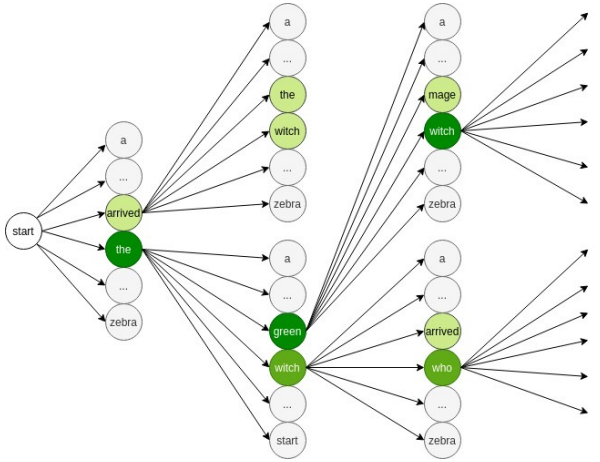
Entrainer un transformer (e.g. GPT)



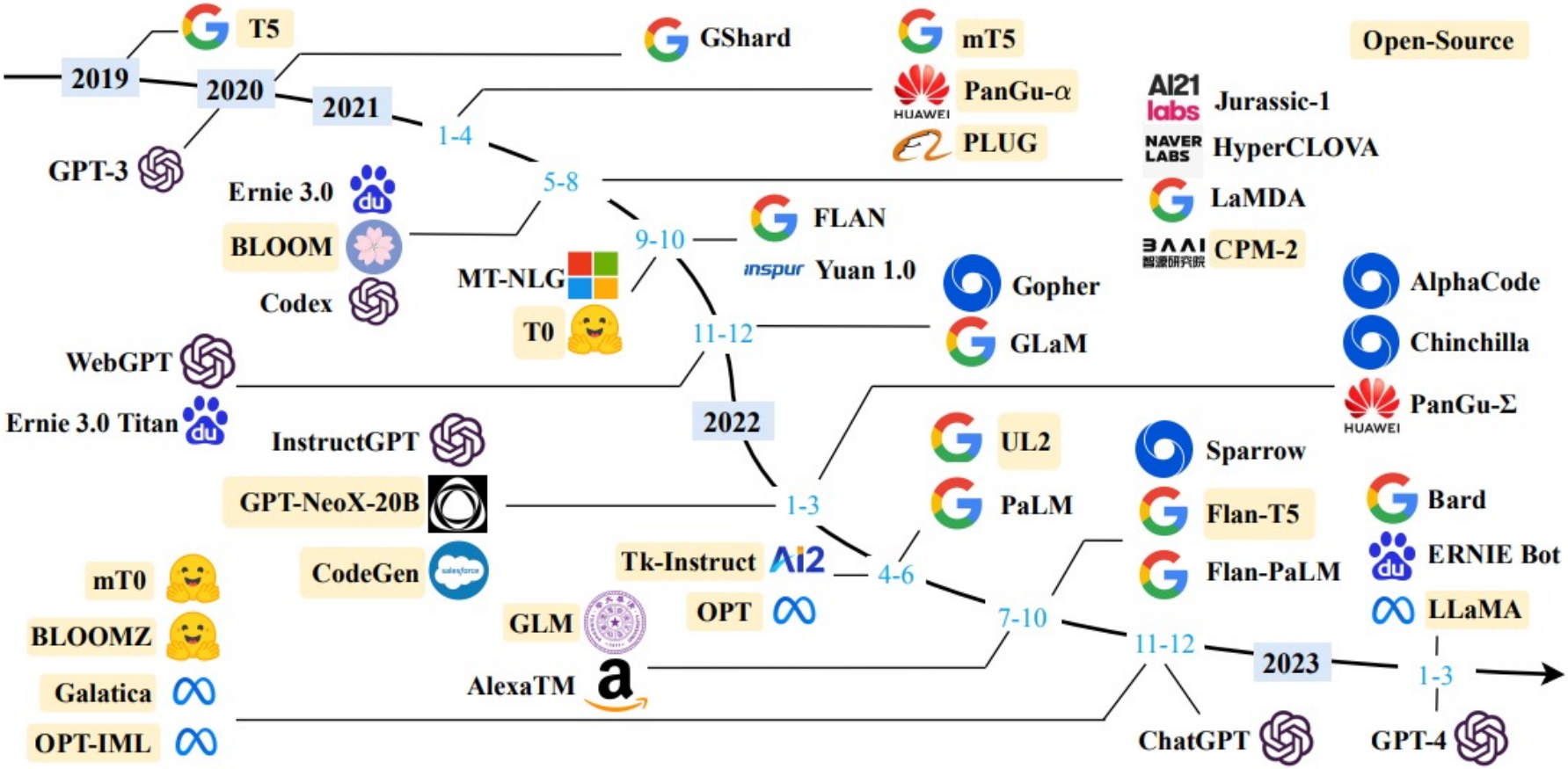
→ Exploiter les architectures précédentes pour écrire du texte



- Génération mot à mot
- Coût très important
- Génération de faisceaux

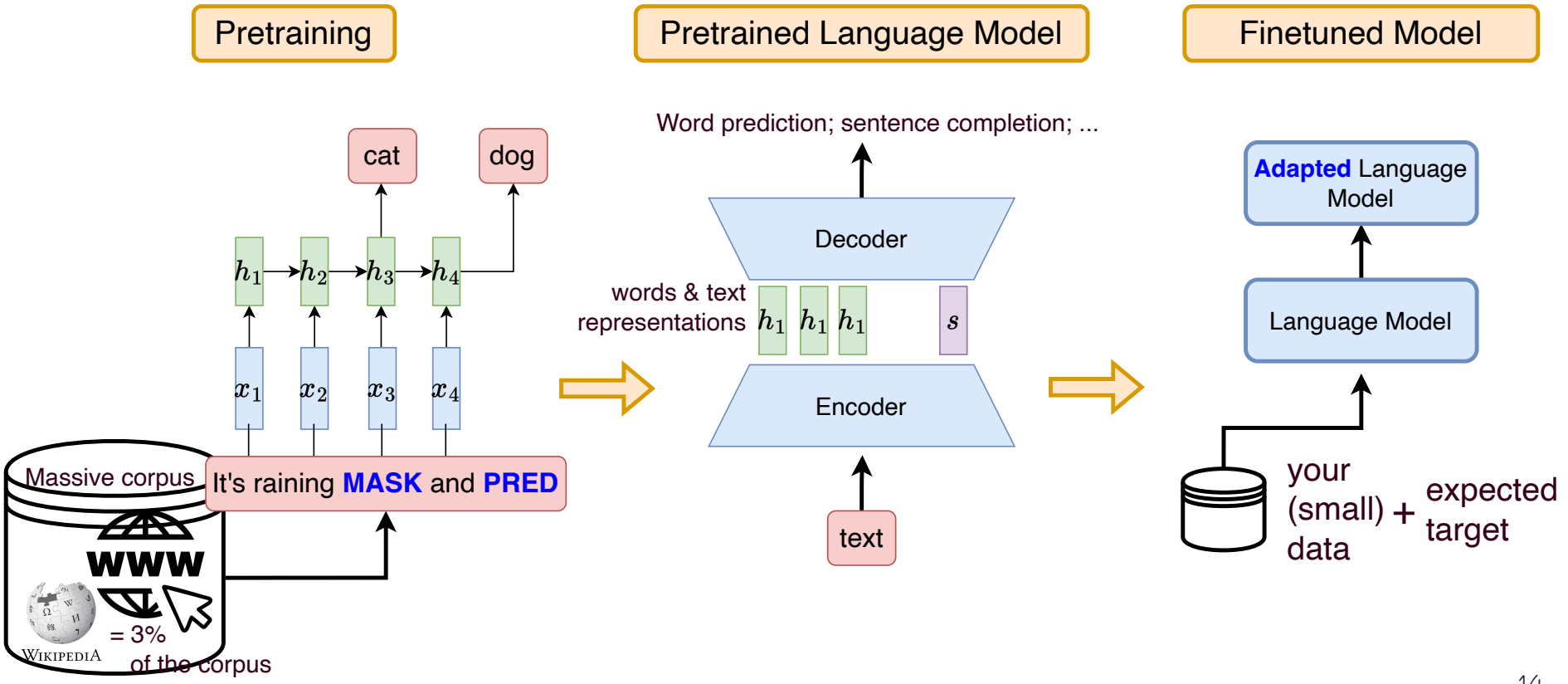


De nombreux modèles de langue



Nouveau paradigme

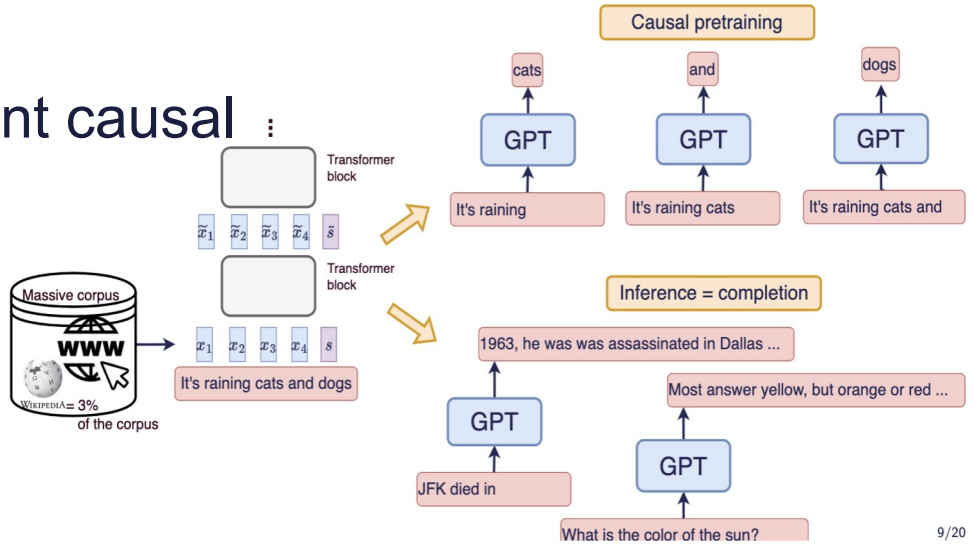
1. Nouvelle tâche
   - ➜ Peu de données
   - ➜ Choix de la taille des modèles
2. Modèle de langue
   - ➜ Connaissances générales
3. Adaptation pour une tâche
   - ➜ Traduction
   - ➜ Détection d'entités nommées
   - ➜ …



Entraînement de ChatGPT

SCIENCES SORBONNE UNIVERSITÉ

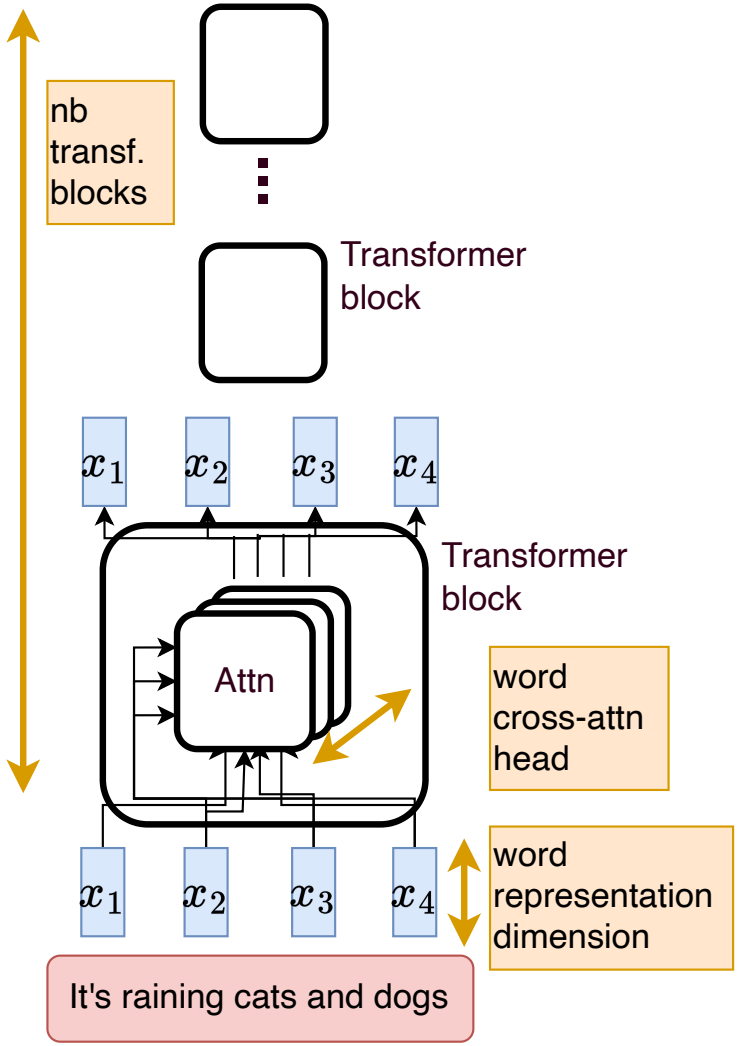
→ Etape 1: Pré-entraînement causal



Un modèle gigantesque

Plus…

- → de mots en entrée [500 =>2k, 32k]
- → de dimensions pour les mots [500-2k =>12k]
- → de têtes d'attention [12 => 96]
- → de blocks/couches [5-12 => 96]



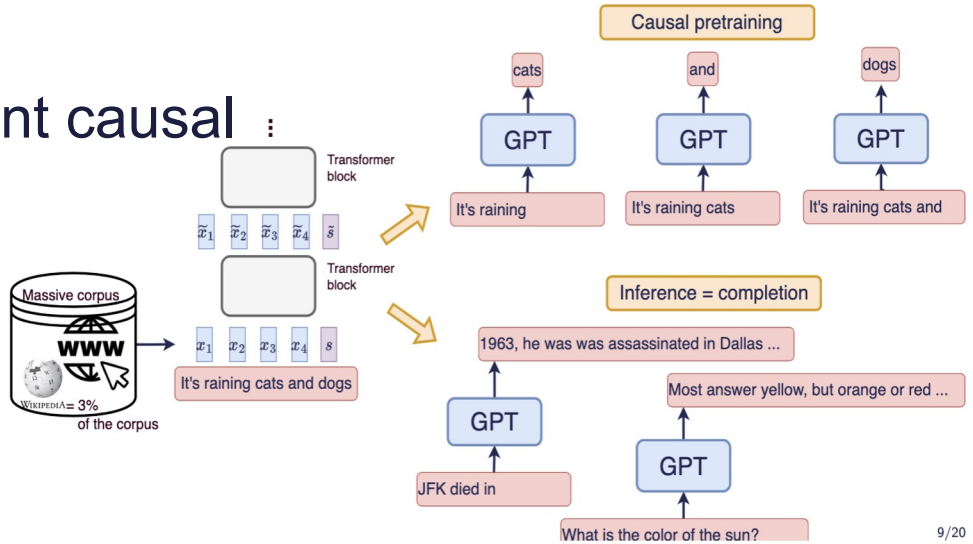
175 Milliards de paramètres... Ca fait quoi?

- $1.75 \cdot 10^{11} \Rightarrow$ 300 Go + 100 Go (stockage des données en inférence) $\approx$ 400Go
- GPU NVidia A100 = 80Go de mémoire (=20k€)
- Coût pour (1) entrainement: 4.6 Millions d'€

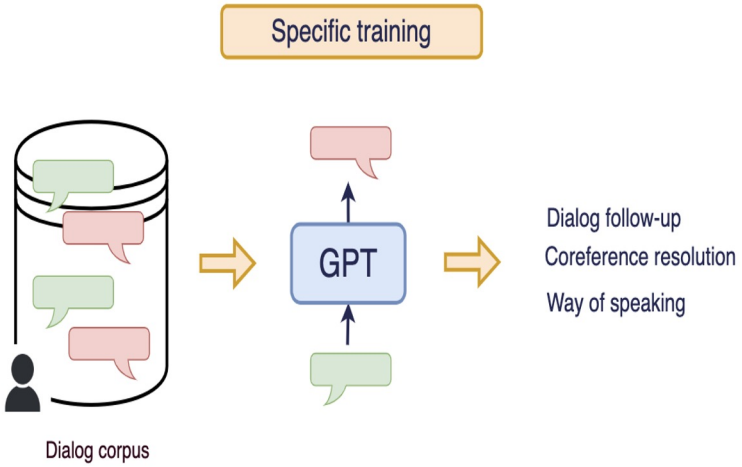
Entraînement de ChatGPT

SCIENCES SORBONNE UNIVERSITÉ

→ Etape 1: Pré-entraînement causal

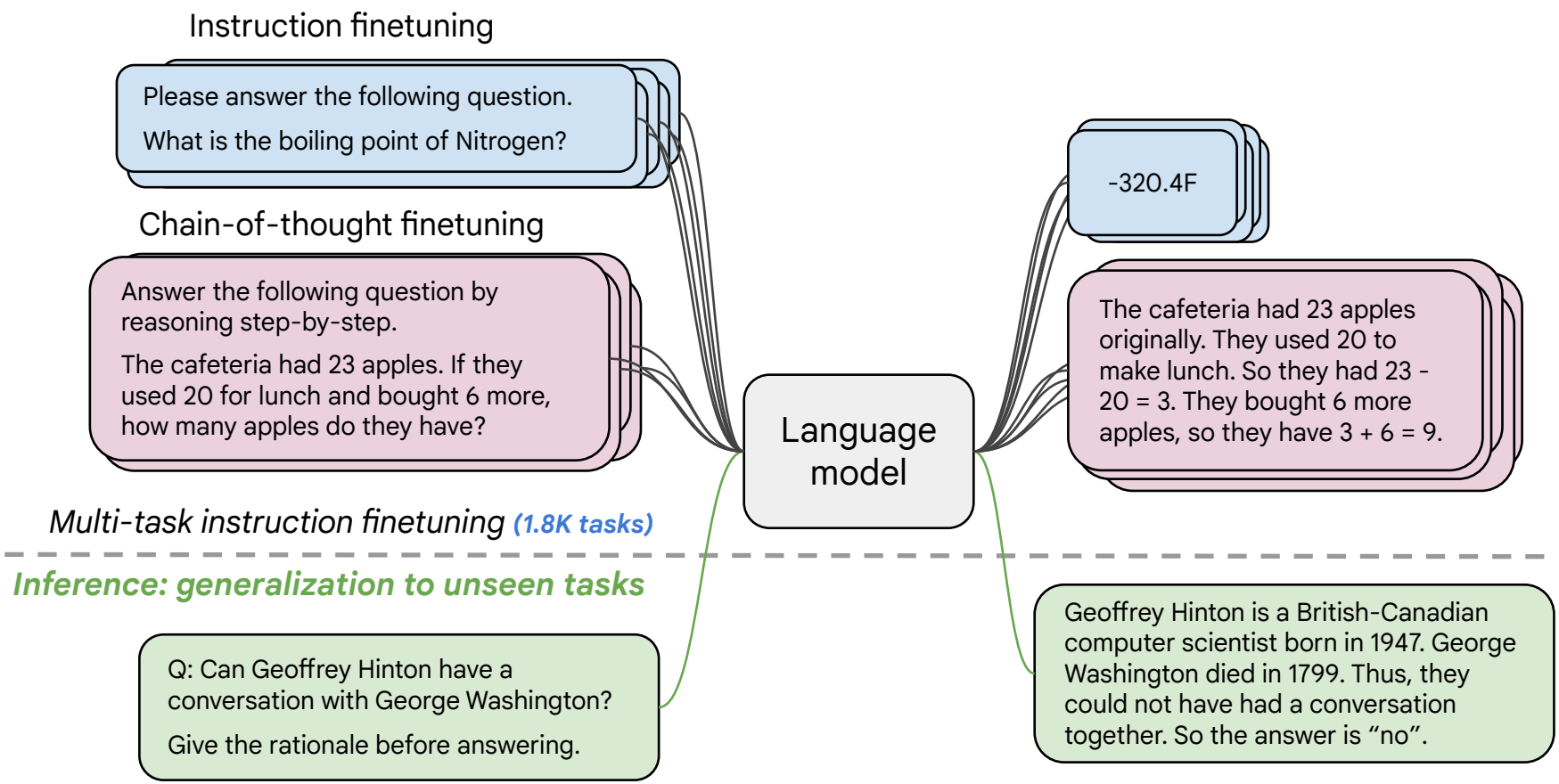


→ Etape 2: Suivi de dialogue



Suivi de dialogue + tâches spécifiques

→ Affinage en questions/réponses, raisonnements, …

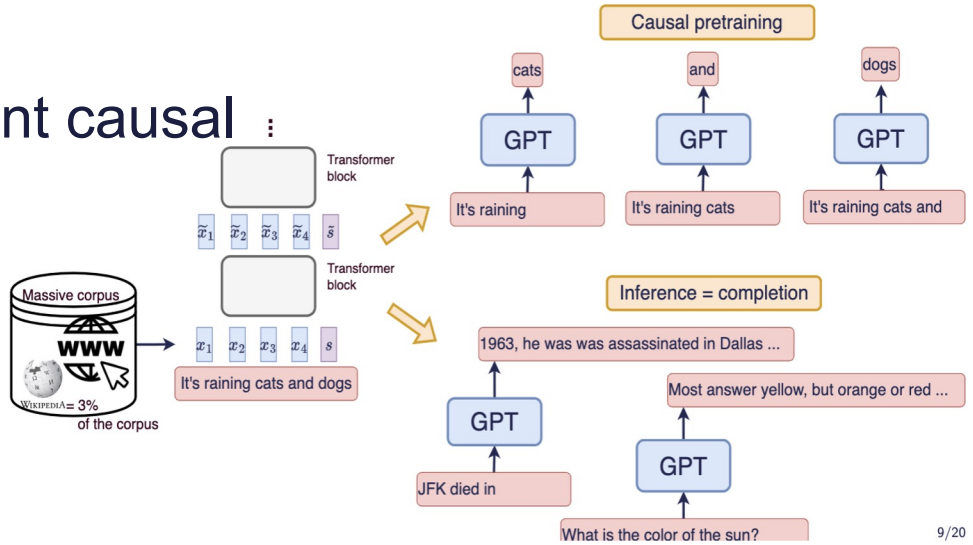


Importance du prompt

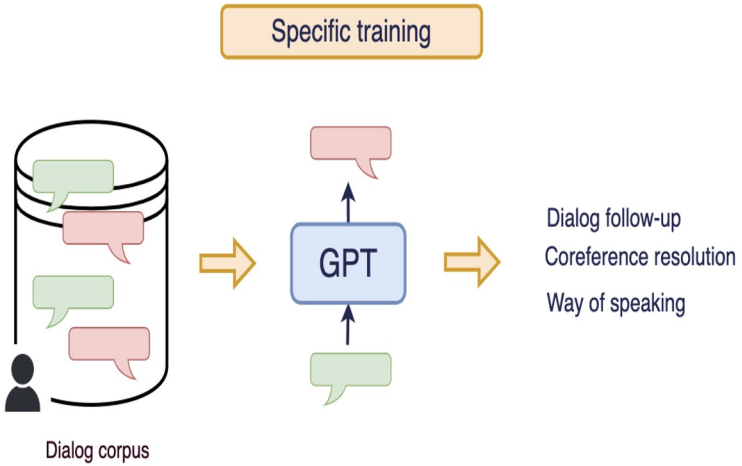
Entraînement de ChatGPT

SCIENCES SORBONNE UNIVERSITÉ

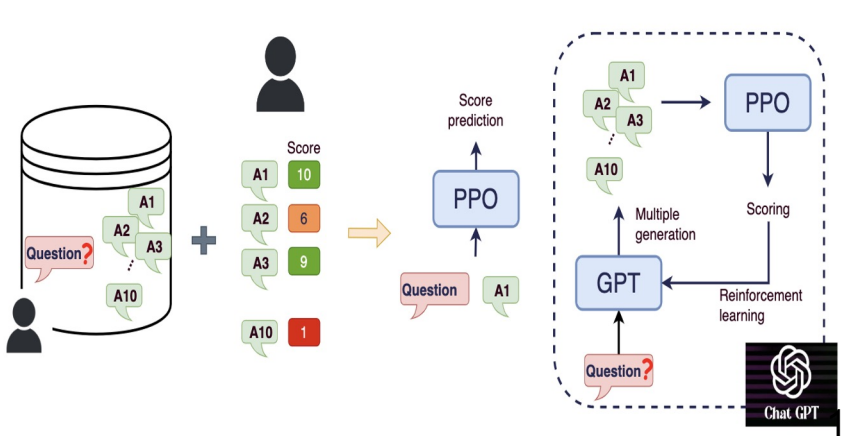
→ Etape 1: Pré-entraînement causal



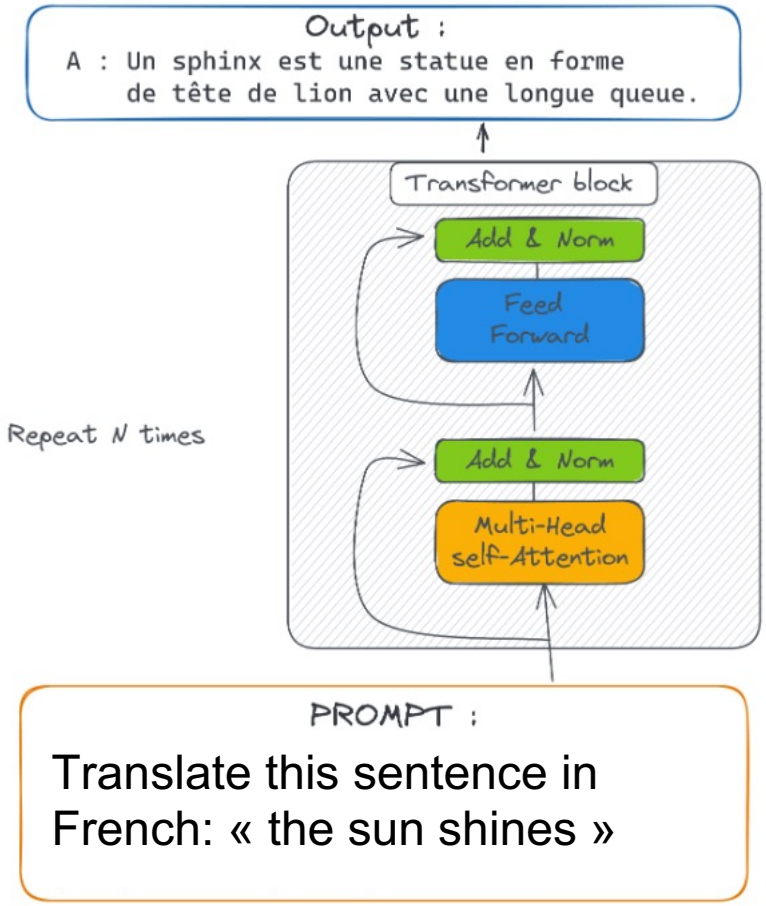
→ Etape 2: Suivi de dialogue



→ Etape 3: Alignement des réponses avec le feedback humain (RLHF)



Le prompt



© Thomas Gerald - 2023

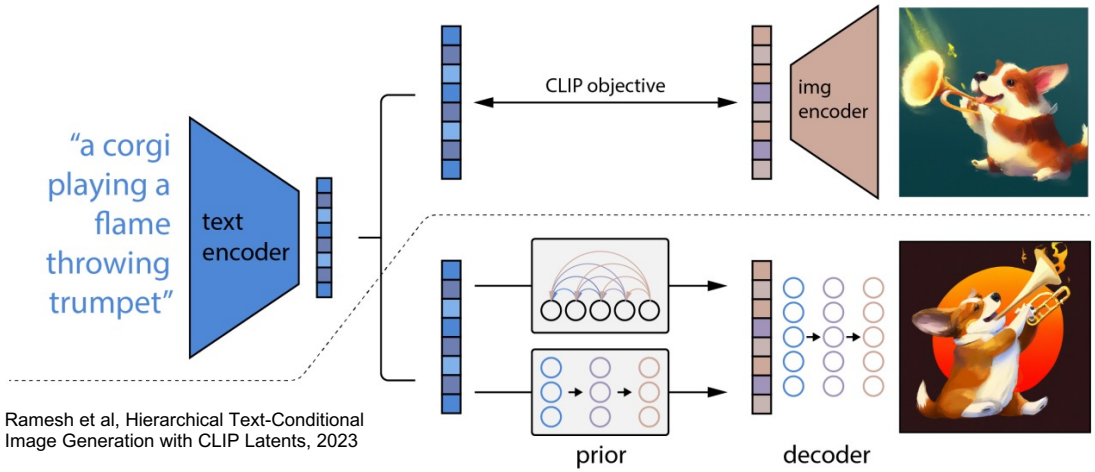
- ➜ Prompt : instruction donnée au modèle de langue
- ➜ Permet d'expliciter la tâche
- ➜ Enjeu : rédiger un prompt précis (contexte, public visé, …)
- ➜ Tout devient génération

Modèles de langue : propriétés intéressantes

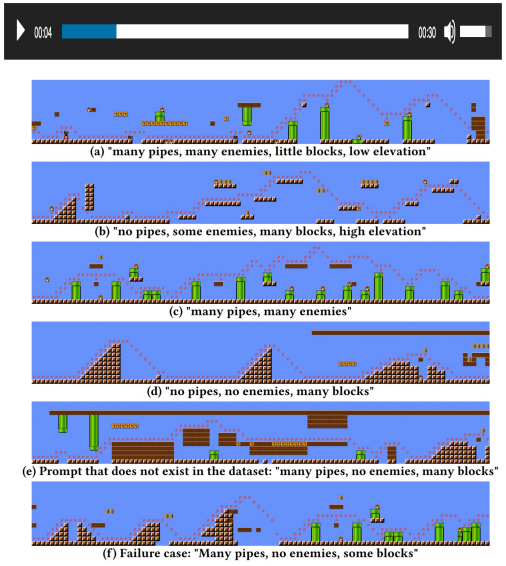
→ Apprentissage en context / In-context learning
- Mentionner des exemples dans le prompt

Multimodal few-shot learning with frozen language models, Tsimpoukelli et al. 2021

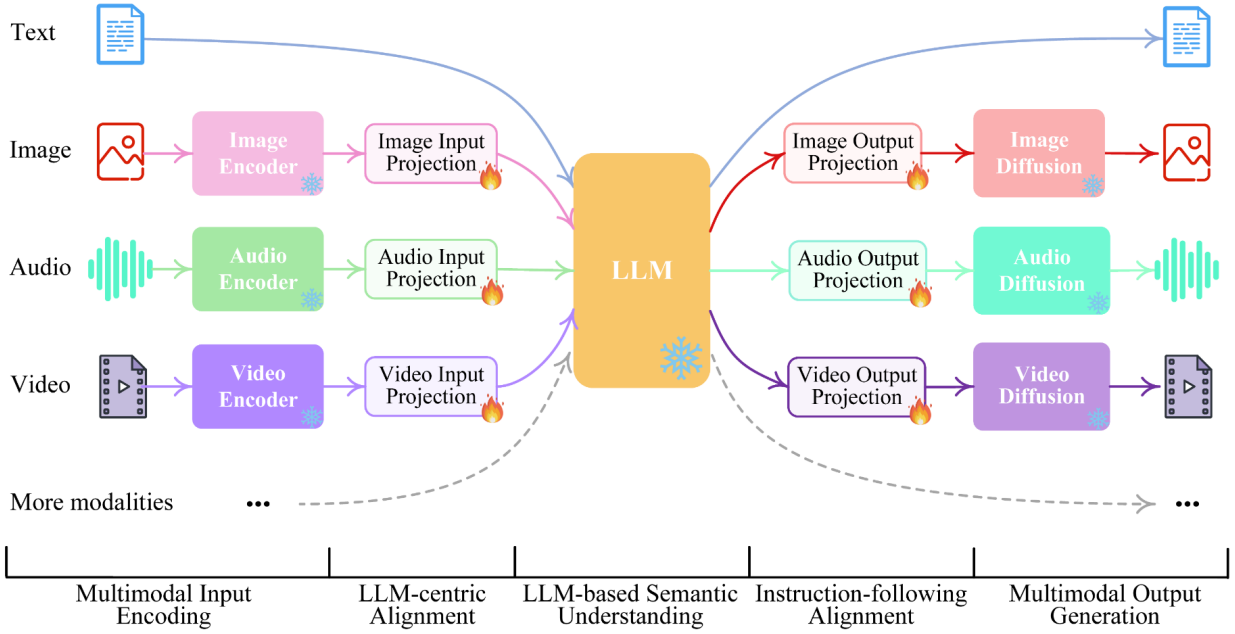
IA générative et multimodalité



Ramesh et al, Hierarchical Text-Conditional Image Generation with CLIP Latents, 2023



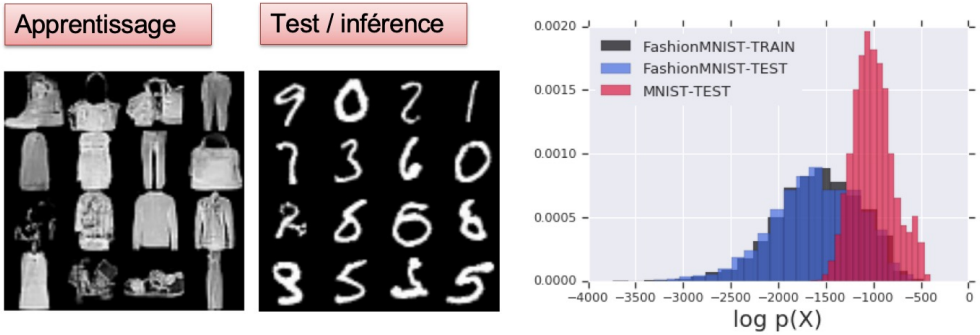
© Midjourney, (Sudhakaran et al 2023)
https://www.blogdumoderateur.com/musiclm-nouvel-outil-ia-google-generer-musique/



Wu et al, NExT-GPT: Any-to-Any Multimodal Large Language Model, 2023

Limites générales

- → Manque de **véracité / fiabilité**
  - → Vérité *vs* Vraisemblance - Génération d'*hallucinations*
- → Incapacité à **s'auto-évaluer**
- → **Biais des données**
- → Manque de **stabilité/prédictibilité**
  - → *How old is Obama* VS *how old is Obama?*
- → Manque d'**explicabilité/interprétabilité**
- → Manque de **transparence**



Foundation Model Transparency Index Scores by Major Dimensions of Transparency, 2023

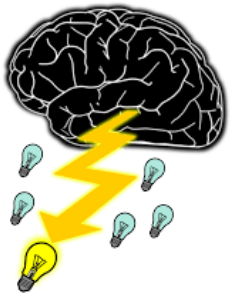
Source: 2023 Foundation Model Transparency Index

Major Dimensions of Transparency	Meta Llama 2	BigScience BLOOMZ	OpenAI GPT-4	stability.ai Stable Diffusion 2	Google PaLM 2	ANTHROPIC Claude 2	cohere Command	AI21labs Jurassic-2	Inflection Inflection-1	amazon Titan Text	Average
Data	40%	60%	20%	40%	20%	0%	20%	0%	0%	0%	20%
Labor	29%	86%	14%	14%	0%	29%	0%	0%	0%	0%	17%
Compute	57%	14%	14%	57%	14%	0%	14%	0%	0%	0%	17%
Methods	75%	100%	50%	100%	75%	75%	0%	0%	0%	0%	48%
Model Basics	100%	100%	50%	83%	67%	67%	50%	33%	50%	33%	63%
Model Access	100%	100%	67%	100%	33%	33%	67%	33%	0%	33%	57%
Capabilities	60%	80%	100%	40%	80%	80%	60%	60%	40%	20%	62%
Risks	57%	0%	57%	14%	29%	29%	29%	29%	0%	0%	24%
Mitigations	60%	0%	60%	0%	40%	40%	20%	0%	20%	20%	26%
Distribution	71%	71%	57%	71%	71%	57%	57%	43%	43%	43%	59%
Usage Policy	40%	20%	80%	40%	60%	60%	40%	20%	60%	20%	44%
Feedback	33%	33%	33%	33%	33%	33%	33%	33%	33%	0%	30%
Impact	14%	14%	14%	14%	14%	0%	14%	14%	14%	0%	11%
Average	**57%**	**52%**	**47%**	**47%**	**41%**	**39%**	**31%**	**20%**	**20%**	**13%**	

Les usages

Les cas d'usage de l'IA générative

SCIENCES SORBONNE UNIVERSITÉ



→ Brainstorming, compte rendu, rédaction de projet
- Développement argumentaire (et recherche de contradiction)
- Mettre en forme des idées
- Reformulation de paragraphes

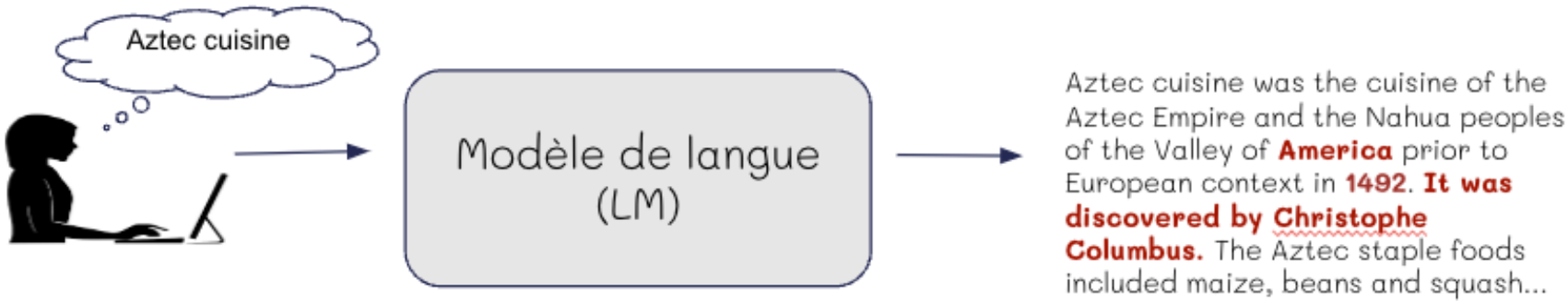
→ Assistant pour le développement informatique
- Génération de code, recherche d'erreurs, …

→ Assistant personnel
- Courrier standard, lettres de recommandation, de motivation, de résiliation, …

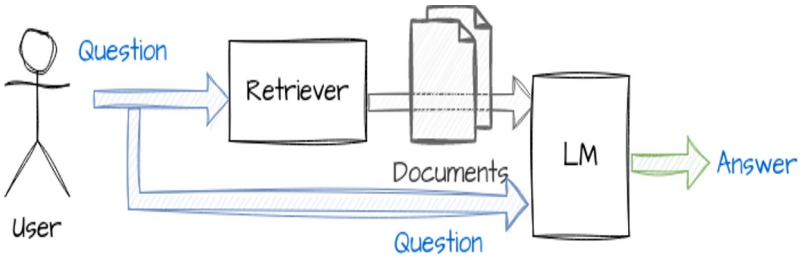
→ Assistant pédagogique
- Wikipédia ++, proposition de plan pour des dissertations, explication de code

→ Analyse de document
- Extraction d'information, question-réponse, …

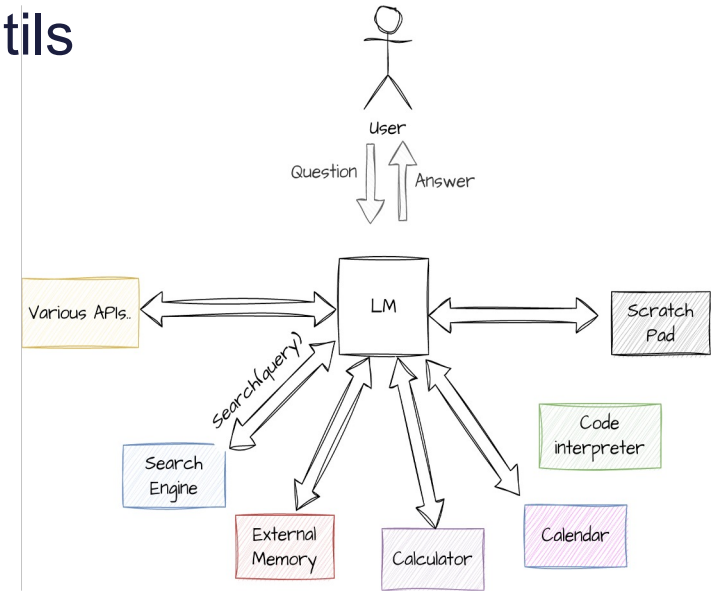
➜ Comment limiter les hallucinations ?



➜ La génération assistée par les outils



Retrieval-augmented generation (RAG)



IA générative : quand l'utiliser ?

- ➜ LLM = mémoire (partielle) d'internet
- ➜ Champion de la reformulation
- ➜ Capacité à comprendre/traduire/générer du code informatique
- ➜ Entrainer à répondre à de nombreux types de questions

- ➜ Oui, ils vont répondre à beaucoup de choses...
  En faisant régulièrement des petites/grosses fautes

Paradigme de la calculatrice:
s'il existe une machine, pourquoi apprendre les tables de multiplication?

Merci !