

Large Language Model

CNRS

21 Novembre 2023

Traffic :

1 million d'utilisateurs en 5 jours

100 millions fin janvier 2023

1.16 milliards en mars 2023

25 millions de visites / jour

12% USA

11% Inde

4% Japon

4% France

3% Canada

2.5% Indonésie

OpenAI :

Valorisation à 29 milliards de dollars

10 milliards d'investissement de Microsoft



Entraînement :

- 160 Milliards de paramètres
- 300 Milliards de mots
- 570 GB de textes

Usage :

Coût : 3 millions de dollars / mois

Revenu prédit : 200 millions de dollars fin 2023

1 milliards de dollars fin 2024

→ Plus de 30 000 cartes graphiques (GPU)

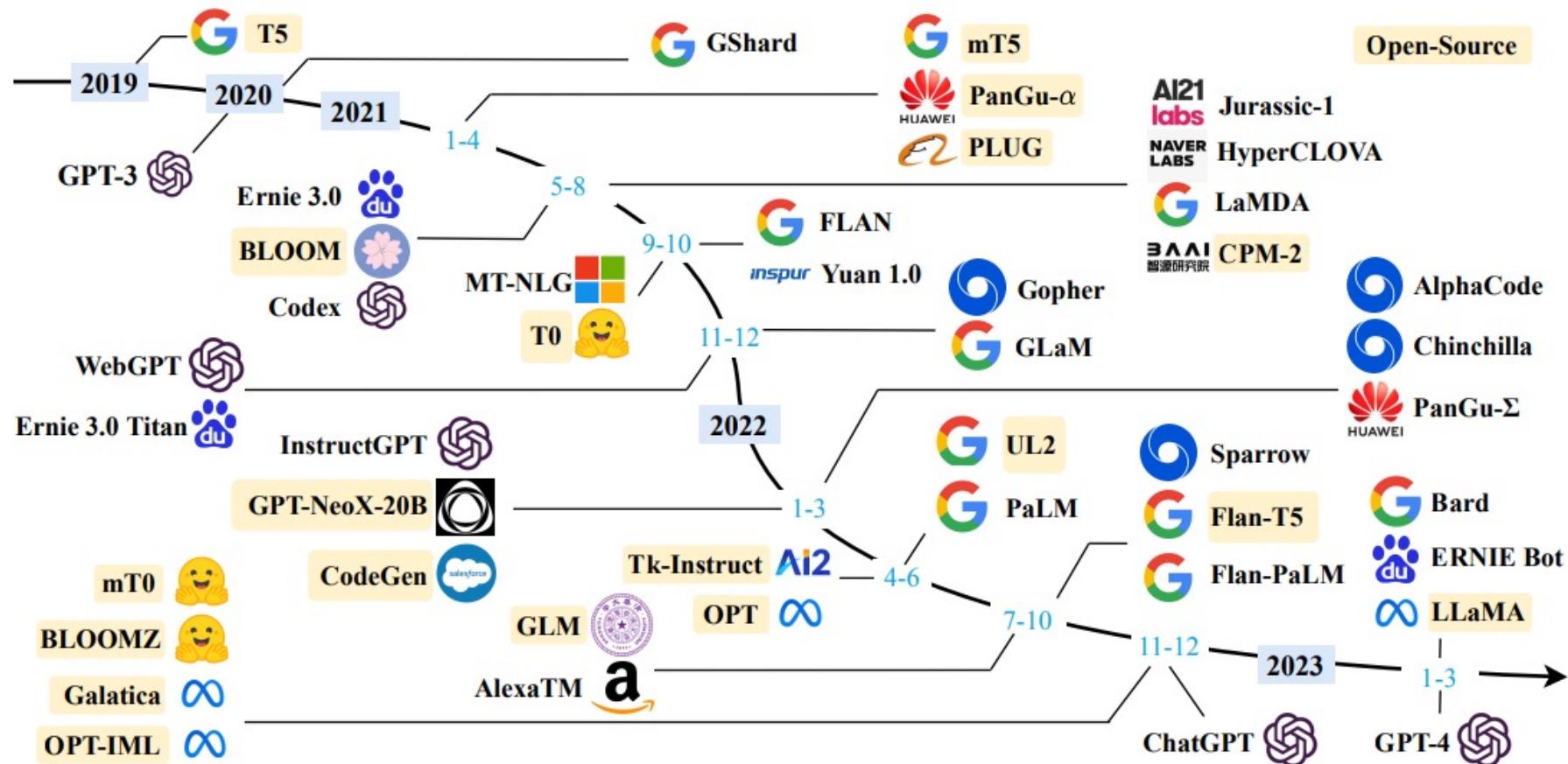


- 315 armoires GPU
- 1 armoire : 0.61m²

Total de 192,15 m²



Mais il n'y a pas que ChatGPT



Modèles dont la taille dépasse 10 Milliards de paramètres

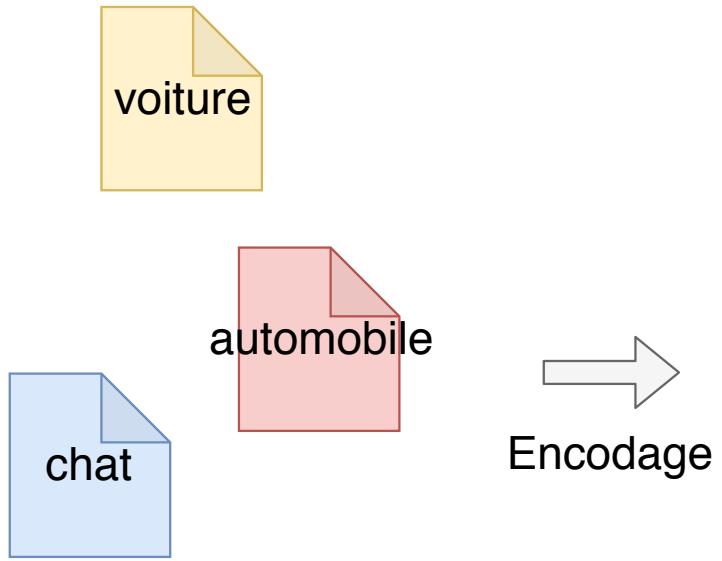
- 1966 ELIZA – psychothérapeute
- Schéma général
 - Question neutre
 - « *Bonjour. Pourquoi venez-vous me voir ?* »
 - Reconnaissance de formes et substitutions dans des phrases prédéfinies
 - « Pourquoi dites-vous **A** ? »
 - Si une question lui était posée, elle demandait pourquoi on lui posait la question
 - Connotation ironique
 - « je comprends »
- SHRDLU (actions dans le monde réel, 1972), PARRY (simulation d'un patient, 1981), FRED (1997), jabberwacky (2005)

```
Welcome to
      EEEEEE  LL      IIII  ZZZZZZ  AAAAAA
      EE     LL      II    ZZ    AA   AA
      EEEE  LL      II    ZZZ  AAAAAAAA
      EE     LL      II    ZZ  AA   AA
      EEEEEE  LLLLLL  IIII  ZZZZZZ  AA   AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU: ■
```

Sémantique vs sacs de mots



Corpus vectorisé

	mot 1	... voiture	... automobile	chat	... mot D
d1	1	0	0	0	0
d2	0	0	1	0	0
d3	0	1	0	0	0

Mêmes
distances

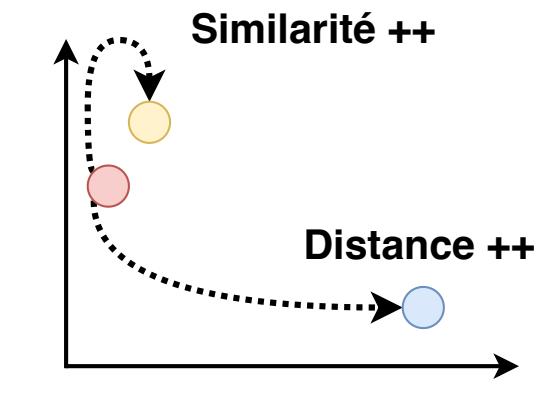
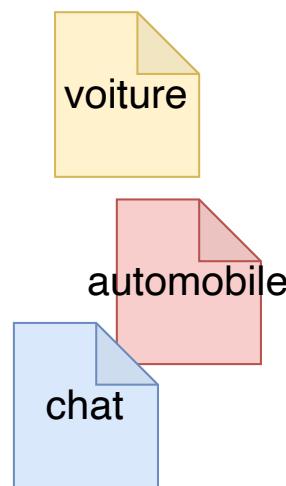
Sémantique = distance entre mots

→ Représentation vectorielle *latente* pour les mots

Corpus en sac de mots

	mot 1	..	voiture	..	automobile	..	chat	..	mot D
d1	1	0	0						
d2	0	0	1						
d3	0	1	0						

Mêmes distances



Comment apprendre efficacement cette représentation?

→ Mot = point dans un espace abstrait (latent)

The fluffy **cat** napped lazily in the sunbeam.

I adopted a stray **cat** from the shelter last week.

My **cat** loves to chase after toy mice.

The black **cat** stealthily crept through the dark alley.

I often find my **cat** perched on the windowsill, watching **birds**.

She gently stroked her **cat**'s fur as it purred contentedly.

Our neighbor's **cat** frequently visits our backyard.

The playful **cat** swatted at the dangling string with its paw.

My **cat** has a preference for fish flavored **cat** food.

The **cat** stealthily stalked a mouse in the garden.

My grandmother has a collection of porcelain **cat** figurines.

The **cat** napped peacefully in the warm sunlight.

The **cat**'s fur was so soft and luxurious to touch.

The **cat**'s eyes gleamed in the darkness as it prowled through the night.

Comprendre = mesurer une distance entre les mots

2012

Apprentissage de représentation,
modèles de langue,
Word2Vec, FastText, ...

Sémantique : comprendre le sens des mots

→ Mot = point dans un espace abstrait (latent)



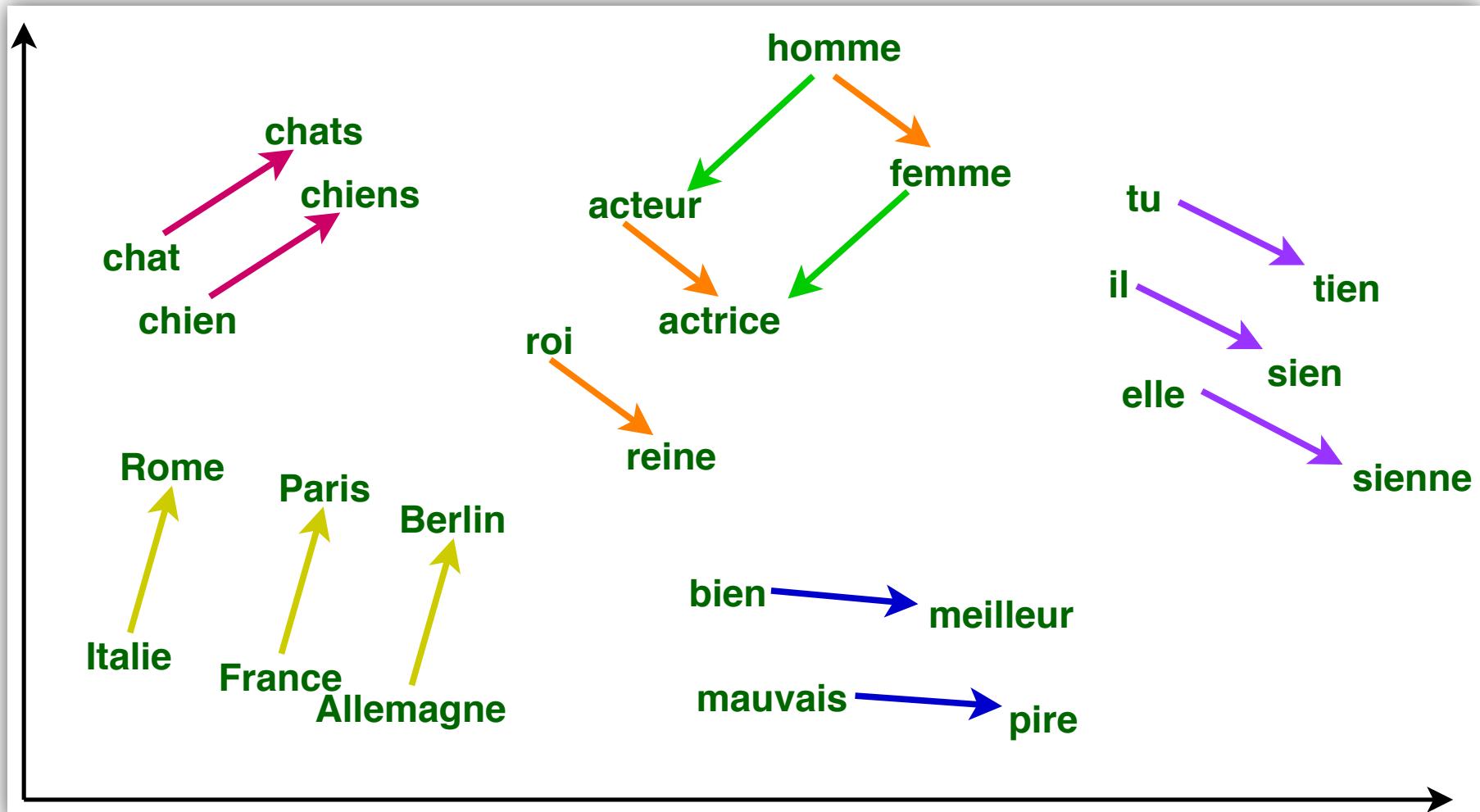
My **dog** is a friendly and playful companion.
The energetic **dog** chased a squirrel up a tree.
The brown **dog** eagerly wagged its **tail** when I came home.
We took our **dog** for a long **walk** in the beautiful park.
The noisy **dog** barked loudly at the passing **mailman**.
The **dog** quickly **fetched** the bright red **ball**.
Our curious **dog** dug a **hole** in the muddy **backyard**.
The friendly **dog** greeted every visitor with a wagging **tail**.
The hungry **dog** eagerly devoured its **dinner**.
The **dog** gently **licked** my hand to show **affection**.
My **sleepy** **dog** curled up on the **comfy** **couch**.
I watched as the **happy** **dog** frolicked in the **sunny** **meadow**.

Comprendre = mesurer une distance entre les mots

2012

Apprentissage de représentation,
modèles de langue,
Word2Vec, FastText, ...

Sémantique : comprendre le sens des mots

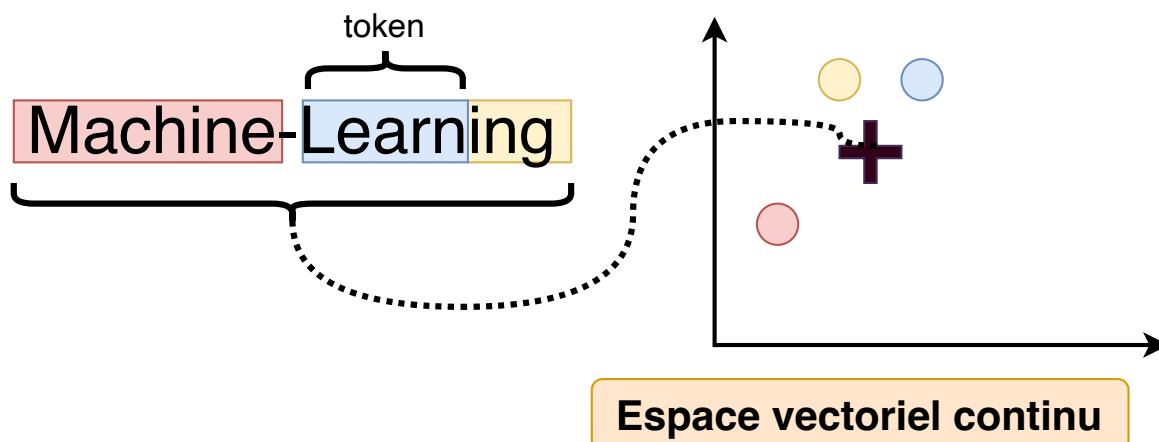


Comprendre = mesurer une distance entre les mots

Word2Vec, FastText, ...

- Il y a toujours de nouveaux mots
 - Fautes, néologismes, vocabulaire technique...

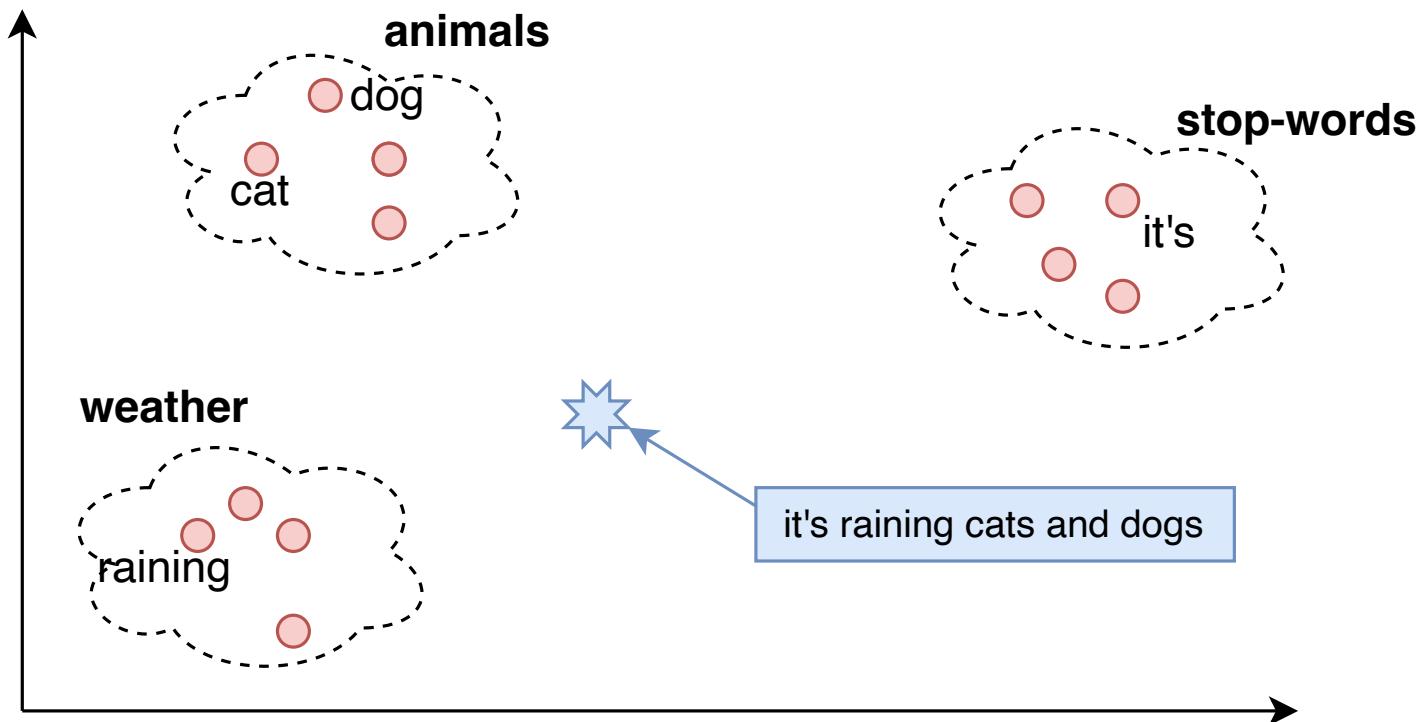
Décomposition en groupes de lettres fréquents



Mots => groupes de lettres universels = token

Changement d'échelle: la phrase

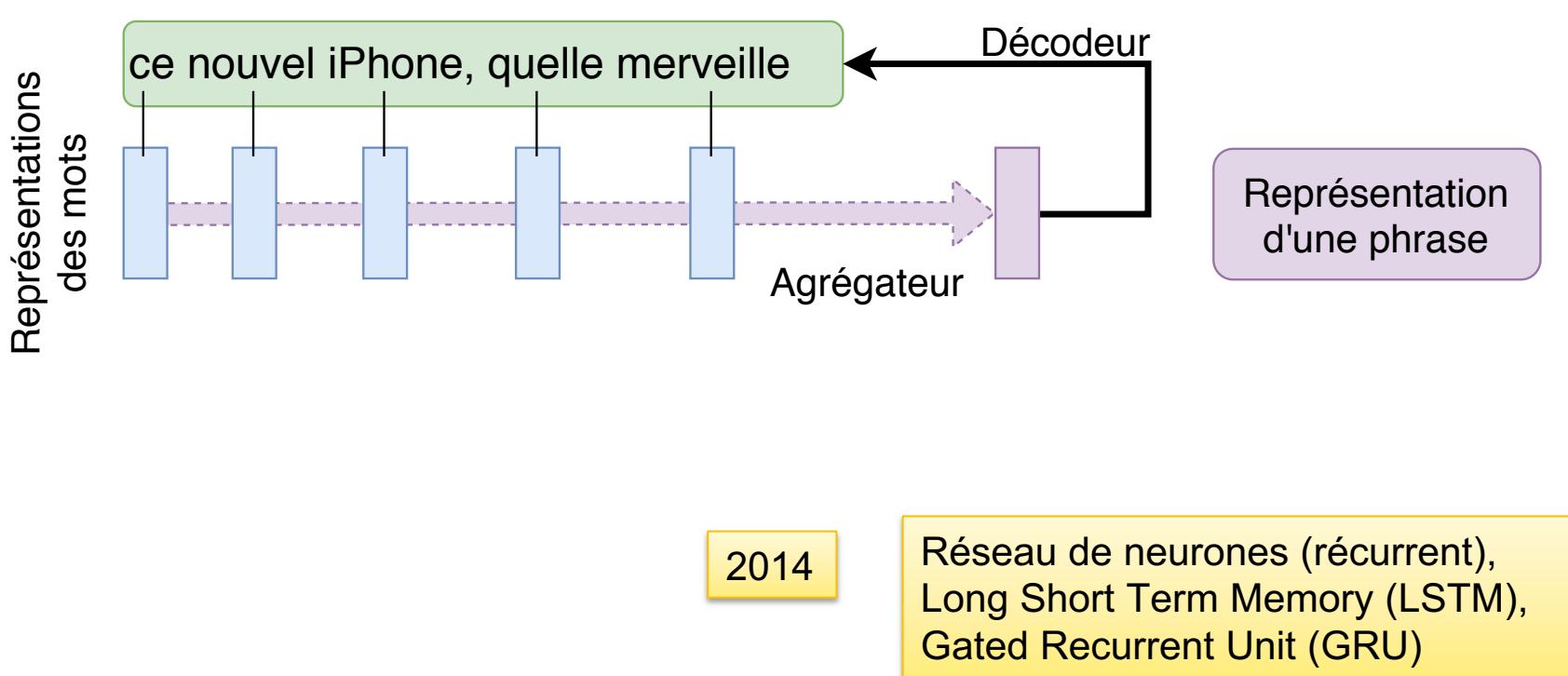
- Approches naïves inefficaces
 - Moyenne des représentations = (souvent) loin de tout



Comprendre les phrases

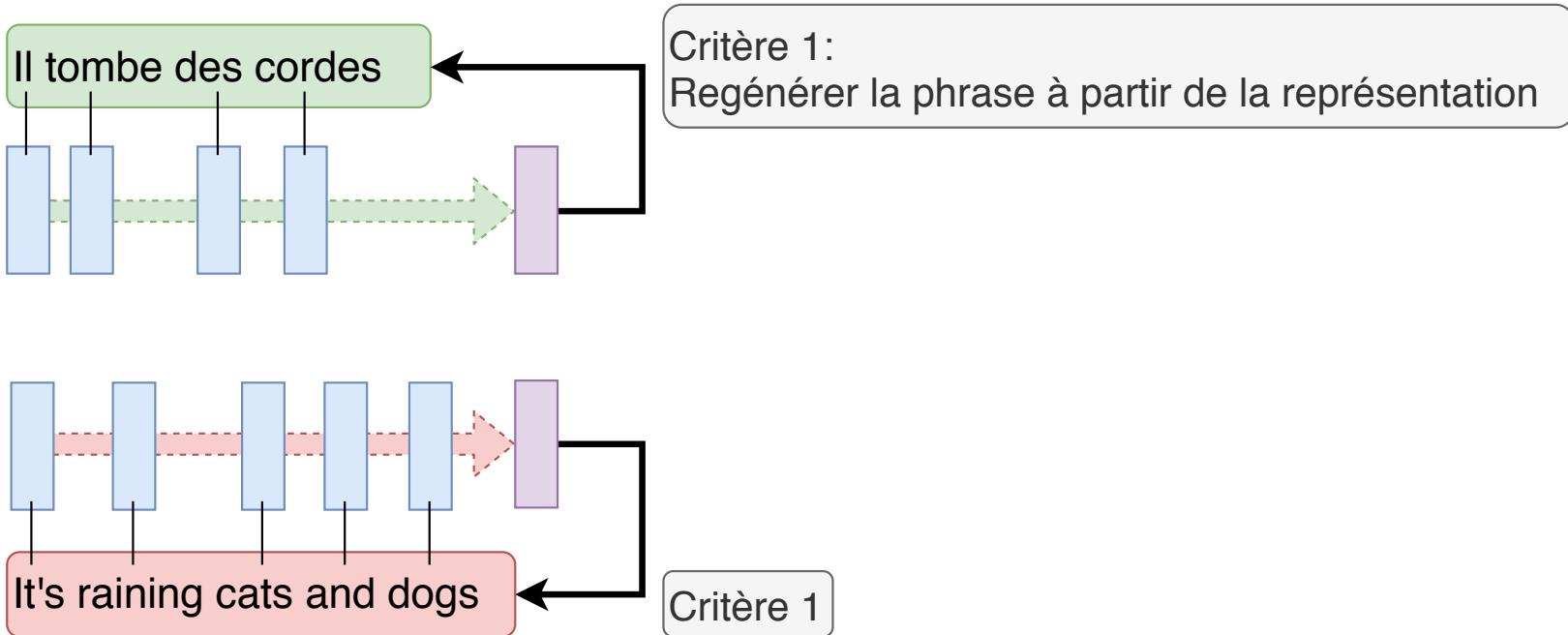
Idée:

- (1) Apprendre à agréger les mots
- (2) Représentation de phrase = capacité à reconstruire le message



→ Une histoire à la mesure de l'IA

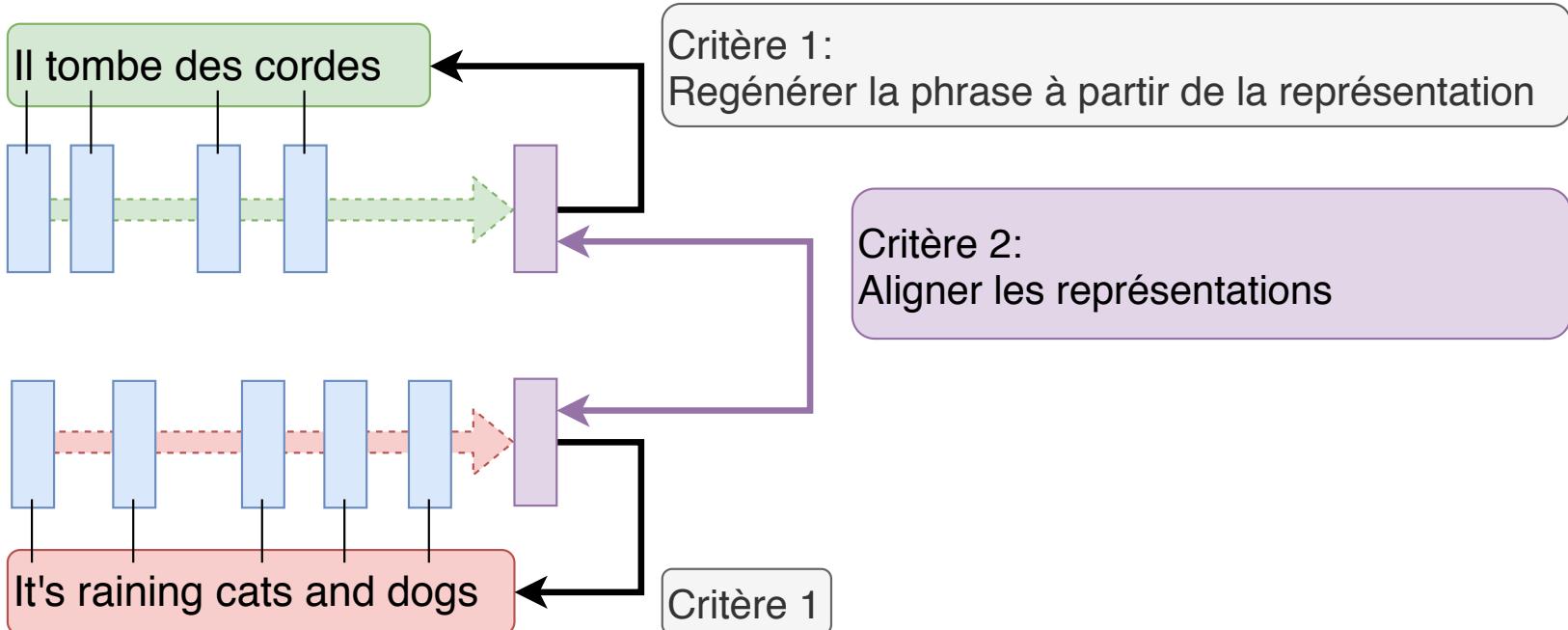
2014



Réseau de neurones récurrents,
LSTM, GRU

→ Une histoire à la mesure de l'IA

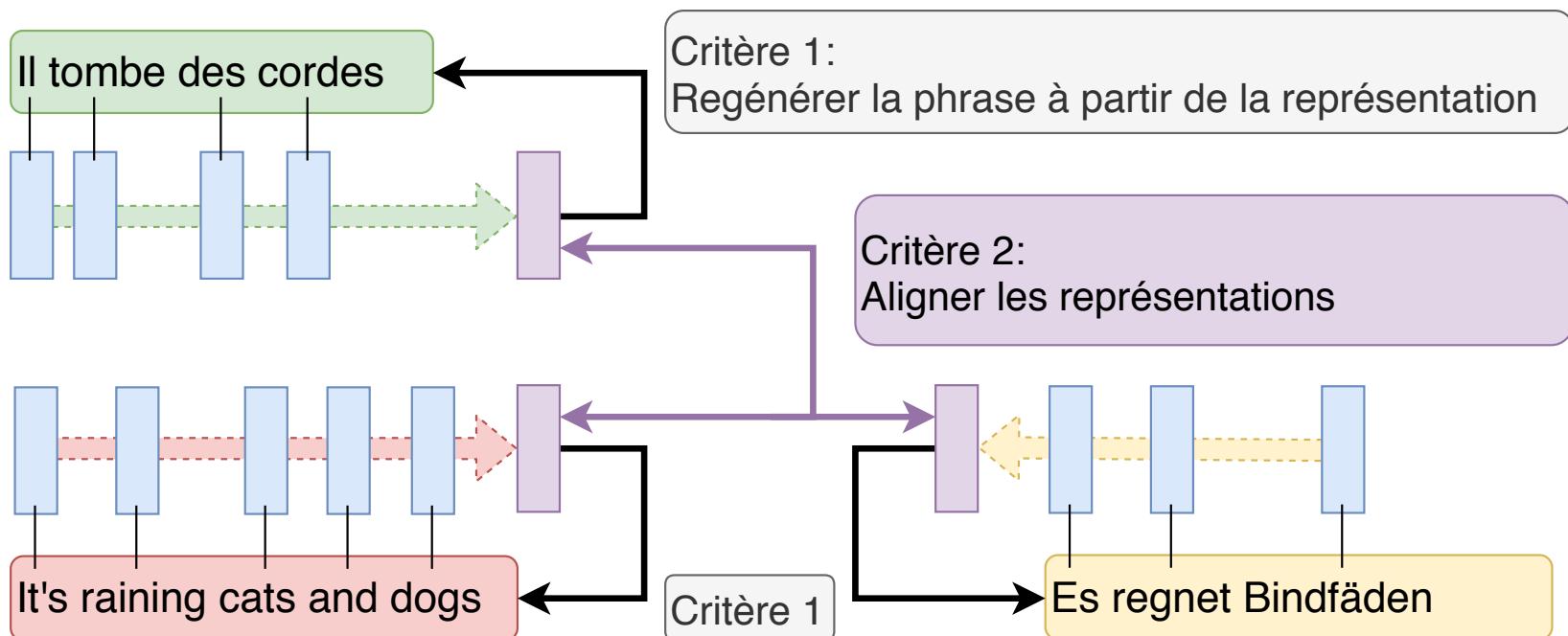
2014



Réseau de neurones récurrents,
LSTM, GRU

→ Une histoire à la mesure de l'IA

2014

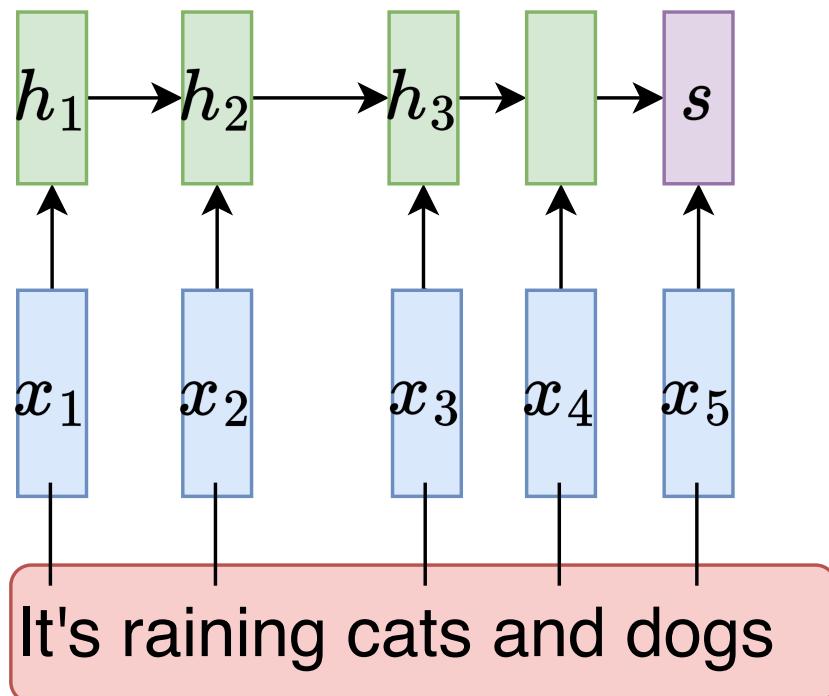


Réseau de neurones récurrents,
LSTM, GRU

Techniques d'agrégation (1)

Réseaux de neurones récurrents

- Agréger récursivement les informations dans h
 - Décision locale ou globale
- Multiples variantes:
 - De gauche à droite + droite à gauche (Bi-RNN)
 - Multi-couches
 - LSTM

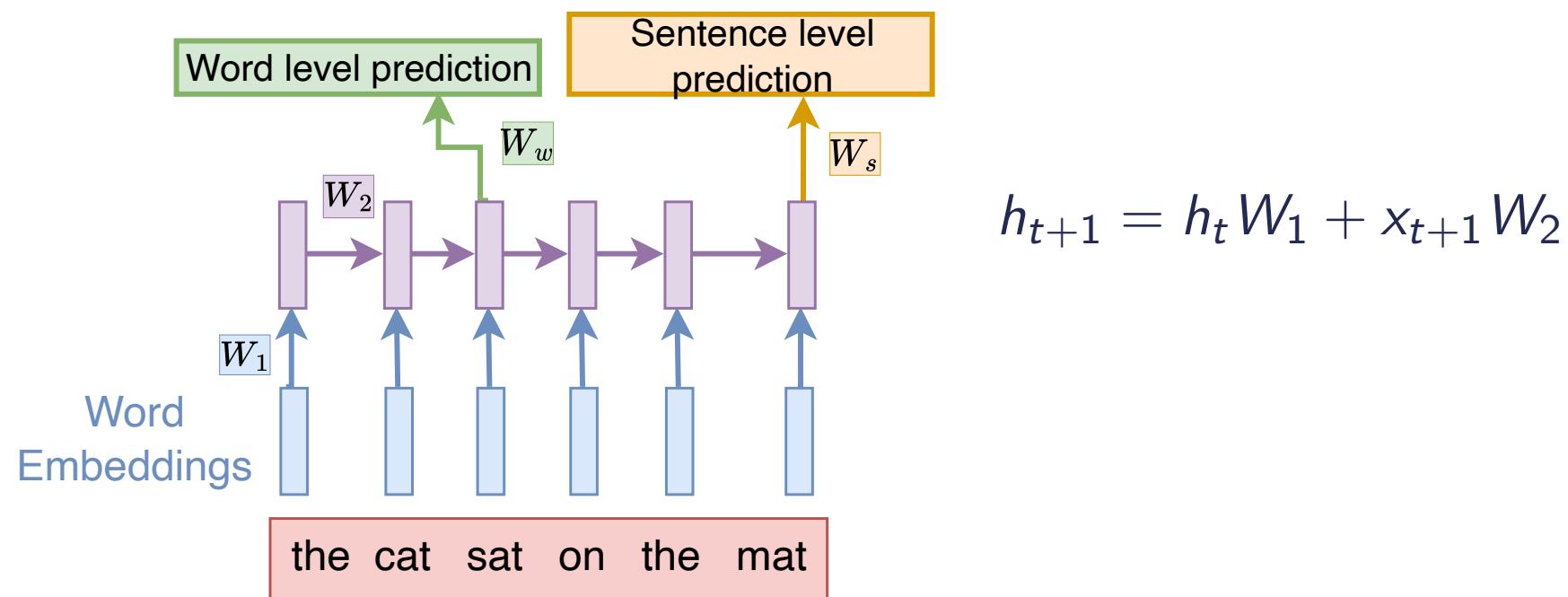


$$h_{t+1} = h_t W_1 + x_{t+1} W_2$$

Techniques d'agrégation (1)

Réseaux de neurones récurrents

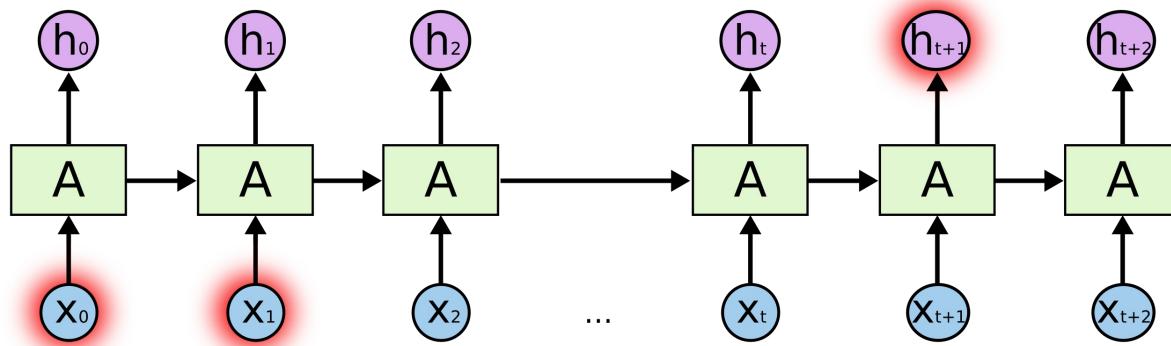
- Agréger récursivement les informations dans h
 - Décision locale ou globale
- Multiples variantes:
 - De gauche à droite + droite à gauche (Bi-RNN)
 - Multi-couches
 - Principe de l'attention : LSTM, GRU, ...



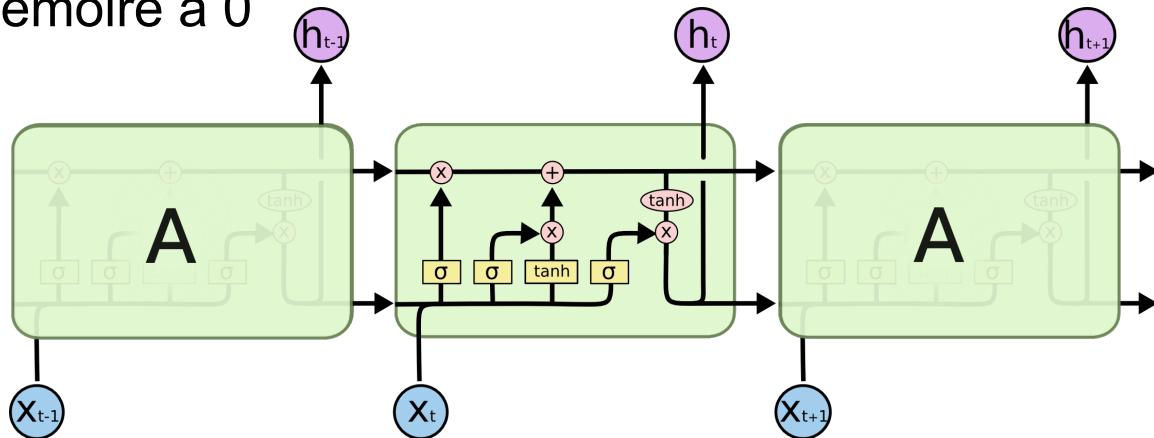
Techniques d'agrégation (1)

Architecture élégante mais instable

- Problème pour la rétro-propagation du gradient dans les longues séquences

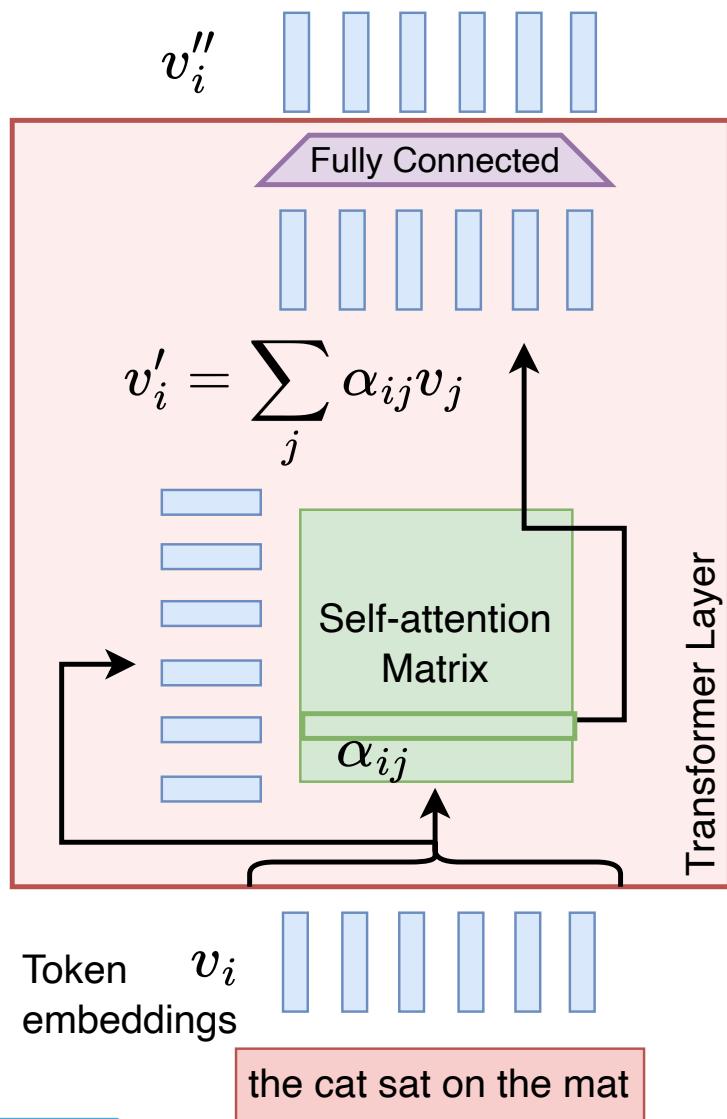


- Architectures à portes (GRU-LSTM)
 - Savoir quand stocker l'information, quand la sauter, quand remettre la mémoire à 0



→ Architecture Transformer = révolution de 2017

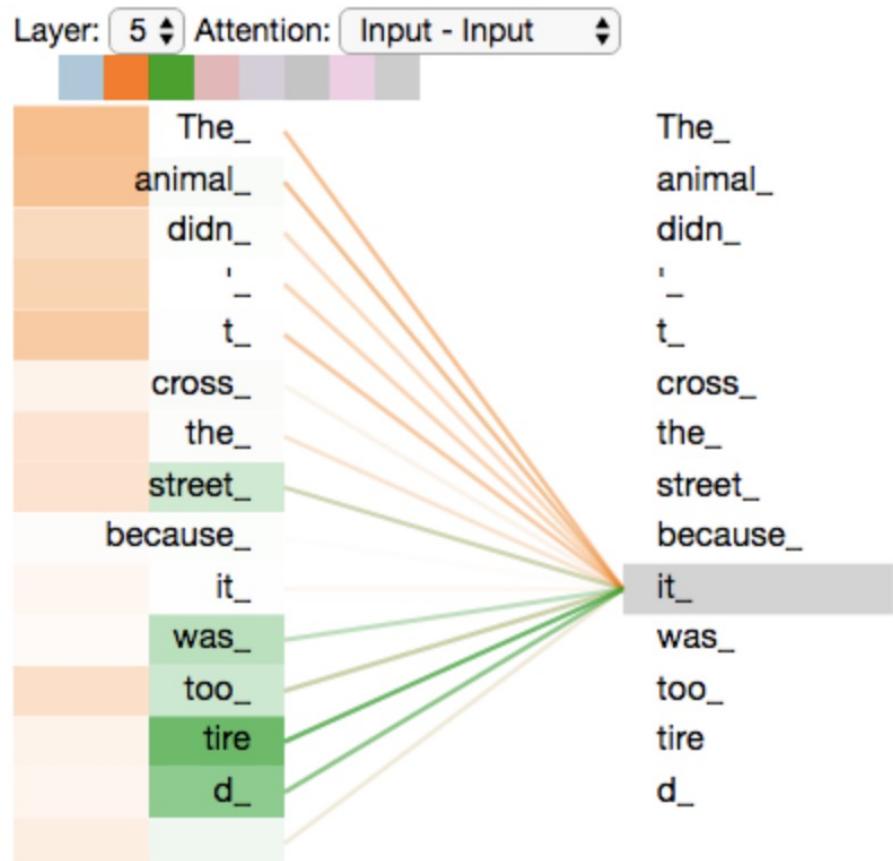
1. Représentation des mots
2. Calcul de l'auto-attention entre les mots
3. Nouvelle représentation des mots prenant en compte la phrase = **représentation contextualisée**



Distinguer : Washington/ville de Washington/homme

→ Architecture Transformer = révolution de 2017

1. Représentation des mots
2. Calcul de l'auto-attention entre les mots
3. Nouvelle représentation des mots prenant en compte la phrase = **représentation contextualisée**

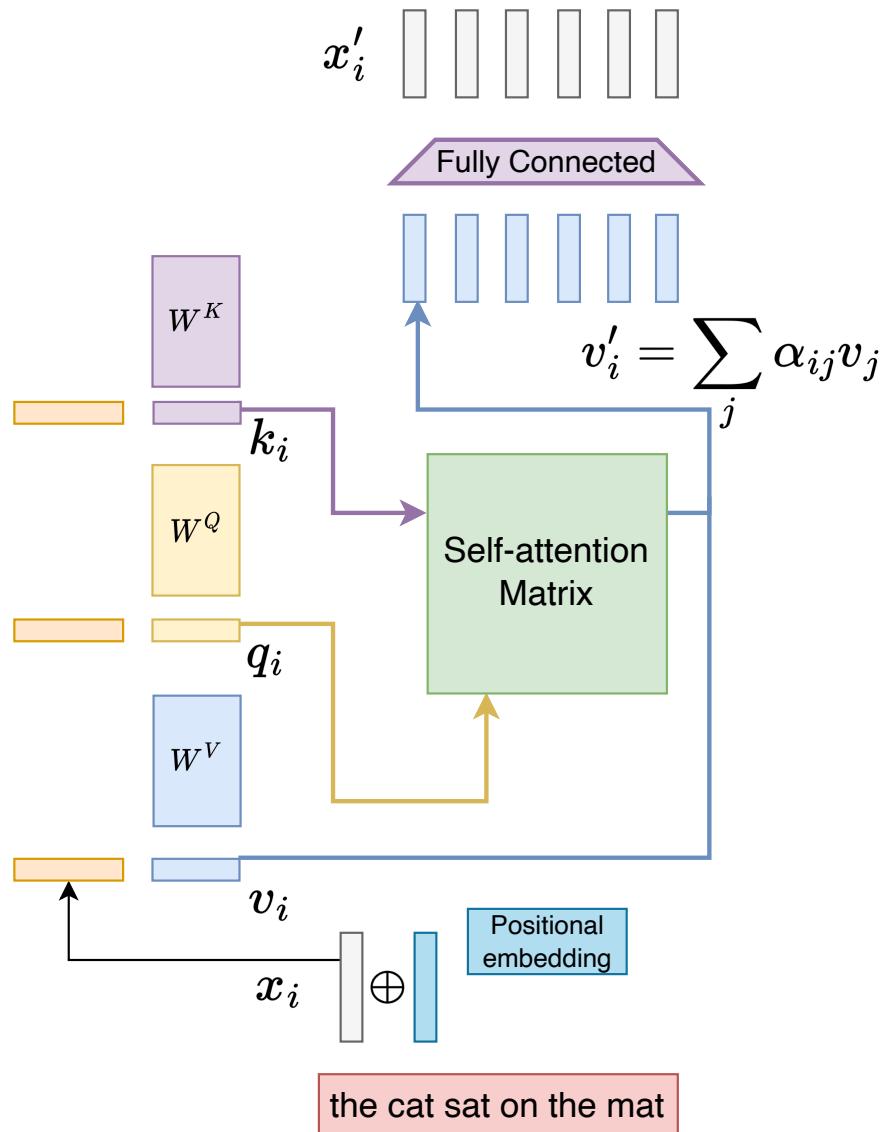


Distinguer : Washington/ville de Washington/homme

→ Architecture Transformer = révolution de 2017

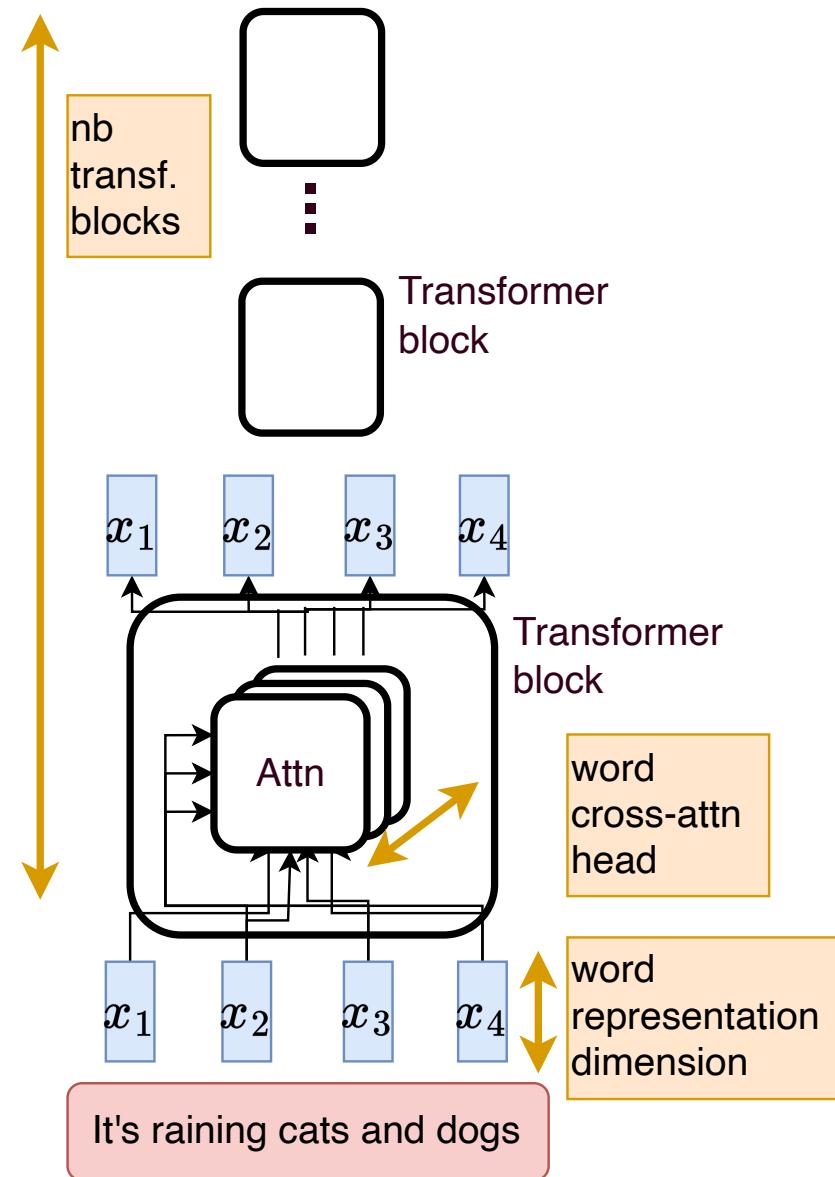
1. Représentation des mots
2. Calcul de l'auto-attention entre les mots
3. Nouvelle représentation des mots prenant en compte la phrase = **représentation contextualisée**

Encoder la position des mots dans
La phrase

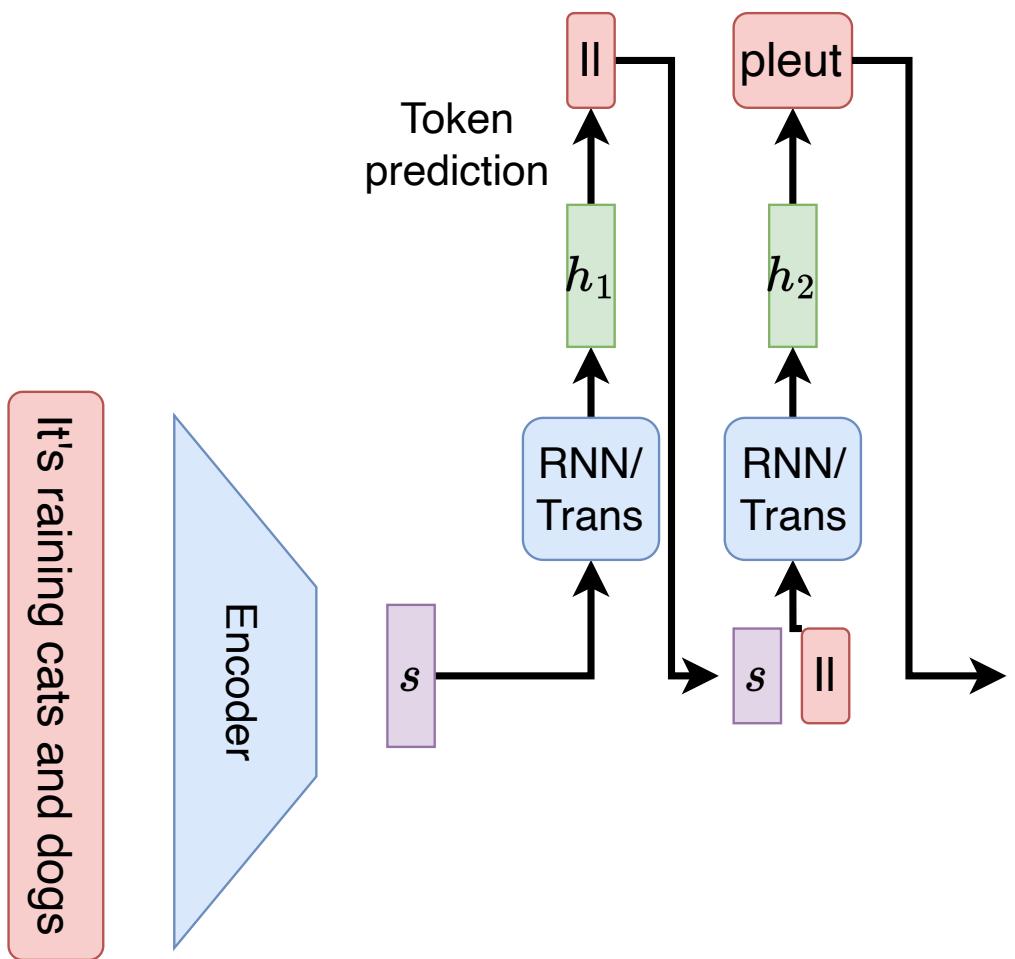


→ Architecture Transformer = révolution de 2017

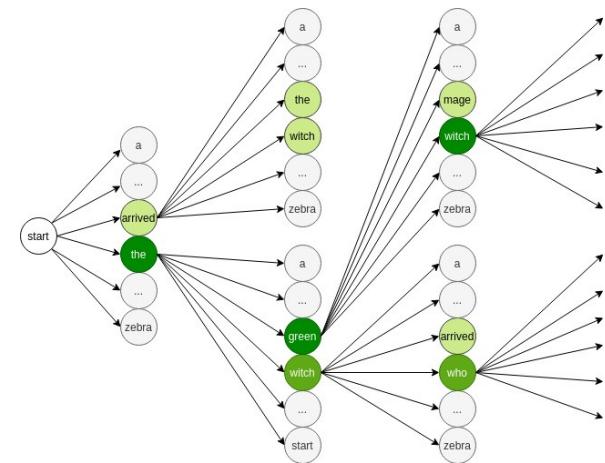
1. Représentation des mots
2. Calcul de l'auto-attention entre les mots
3. Nouvelle représentation des mots prenant en compte la phrase = **représentation contextualisée**
4. Itérer le processus sur de nombreuses couches



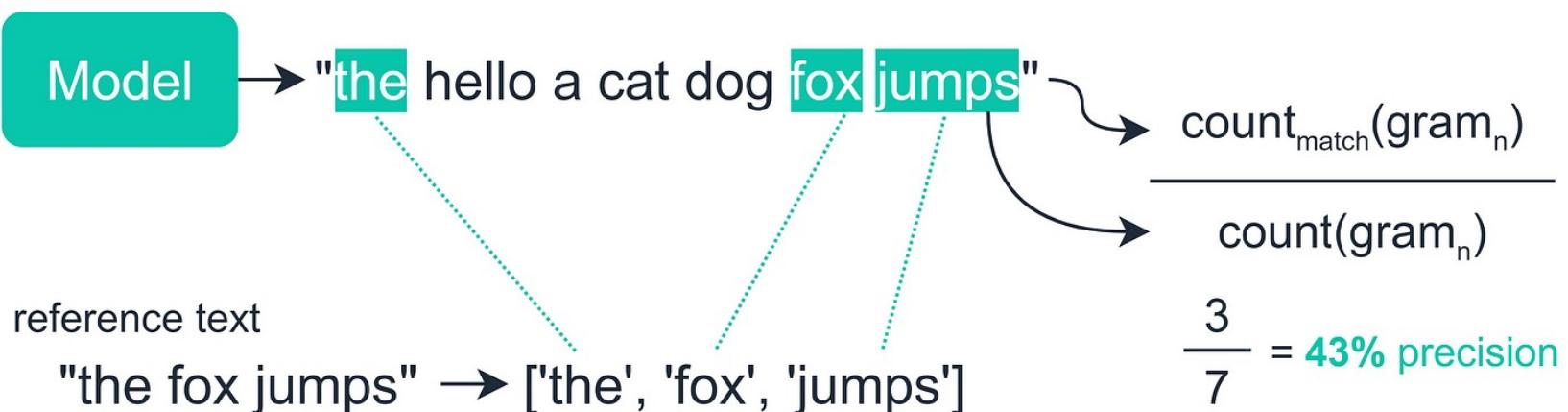
→ Exploiter les architectures précédentes pour écrire du texte



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



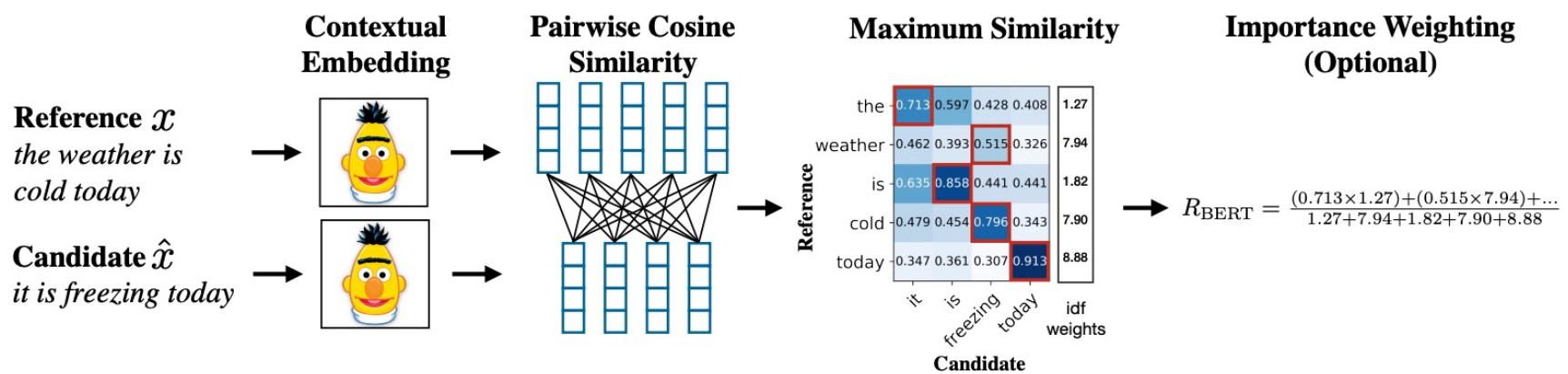
- Multiplier les **applications**: chatbot, résumé automatique, ...
- Défi: **évaluer** ces systèmes pour les faire progresser
 1. Obtenir une vérité terrain
 2. Mesurer un écart entre proposition & vérité terrain



Evaluer quantitativement les textes générés est très difficile

Les métriques d'évaluation pour la génération

- Comparaison sémantique : BertScore
 - Evaluer le sens des phrases/des mots



Les métriques d'évaluation

➤ Evaluation humaine : critique pour la génération

→ Critères :

- Variabilité des styles, du vocabulaire
- Fluidité du texte
- Couverture
- ...

→ Evaluation subjective

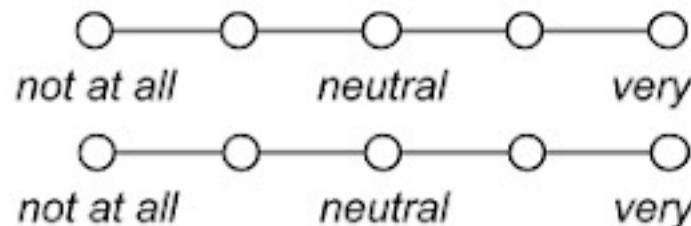
- Plusieurs annotateurs
- Accord inter-annotateurs

Input: Bud Powell était un pianiste de légende.

Reference: Bud Powell was a legendary pianist.

Candidate: Bud Powell was a great pianist.

How fluent is the sentence?



Does it accurately convey the meaning of the reference?

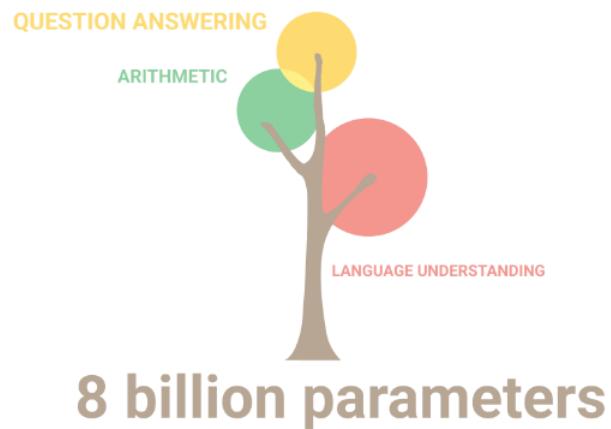
Chapitre 5 :

Modèles de langue

Apprentissage vs. Adaptation

→ Qu'apprend un modèle ?

- Grammaire, conjugaison de verbe, ...
- Connaissances/sémantique
- Mais aussi.... Des métapatrons?



Capacités émergentes : des études empiriques

Emergent Abilities of Large Language Models

Jason Wei¹

Yi Tay¹

Rishi Bommasani²

Colin Raffel³

Barret Zoph¹

Sebastian Borgeaud⁴

Dani Yogatama⁴

Maarten Bosma¹

Denny Zhou¹

Donald Metzler¹

Ed H. Chi¹

Tatsunori Hashimoto²

Oriol Vinyals⁴

Percy Liang²

Jeff Dean¹

William Fedus¹

jasonwei@google.com

yitay@google.com

nlprishi@stanford.edu

craffel@gmail.com

barrettzoph@google.com

sborgeaud@deepmind.com

dyogatama@deepmind.com

bosma@google.com

dennyyzhou@google.com

metzler@google.com

edchi@google.com

tashim@stanford.edu

vinyals@deepmind.com

pliang@stanford.edu

jeff@google.com

liamfedus@google.com

¹ Google Research ² Stanford University ³ UNC Chapel Hill ⁴ DeepMind

Are Emergent Abilities in Large Language Models just In-Context Learning?

Sheng Lu^{1*}, Irina Bigoulaeva^{1*}, Rachneet Sachdeva¹,
Harish Tayyar Madabushi², and Iryna Gurevych¹

¹ Ubiquitous Knowledge Processing (UKP) Lab, Technische Universität Darmstadt

² Department of Computer Science, The University of Bath

www.ukp.tu-darmstadt.de

htm43@bath.ac.uk

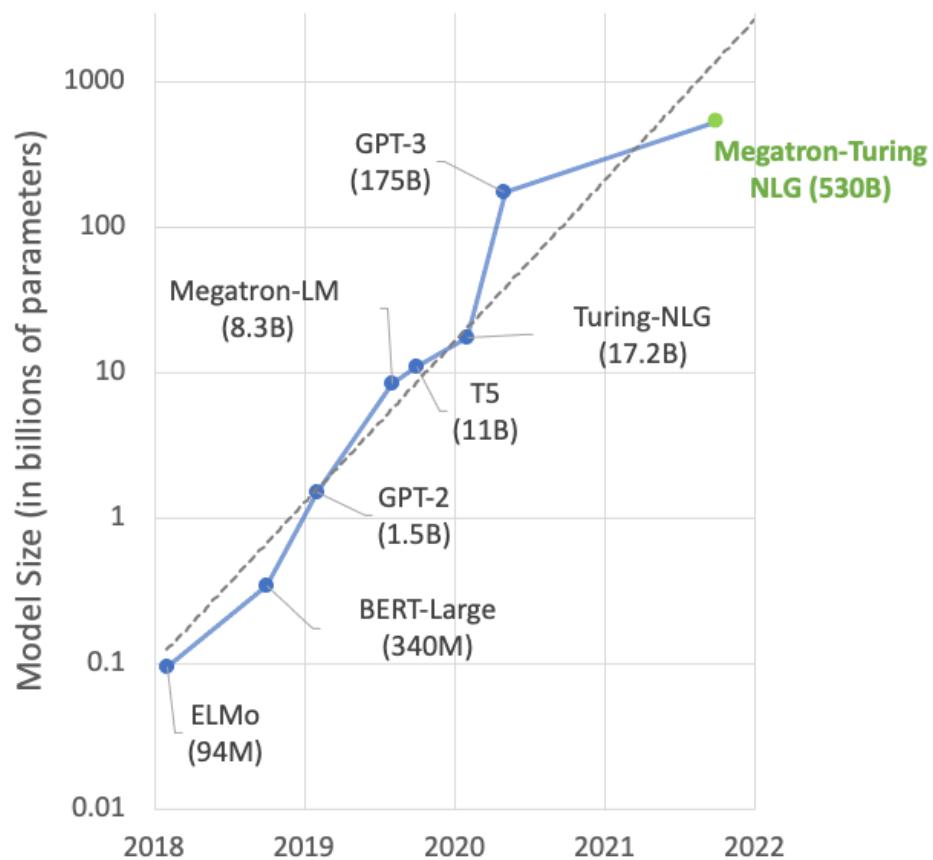
Family	Model	Instruction-tuned	Size
GPT	GPT-2	GPT-2-IT*	117M
	GPT-2-XL	GPT-2-XL-IT*	1.6B
	GPT-J	GPT-JT	6.7B
	davinci	text-davinci-001 text-davinci-003	175B
T5	T5-small	FLAN-T5-small	60M
	T5-large	FLAN-T5-large	770M
Falcon	Falcon-7B	Falcon-7B-IT	7B
	Falcon-40B	Falcon-40B-IT	40B
LLaMA	LLaMA-7B	–	7B
	LLaMA-13B	–	13B
	LLaMA-30B	–	35B

Capacités émergentes : des études empiriques

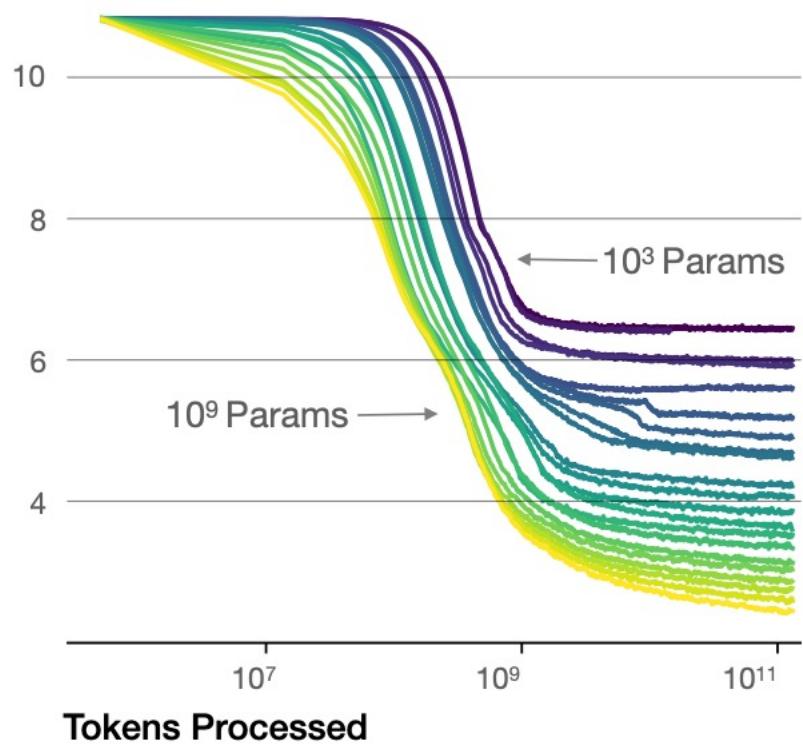
Table 1: List of emergent abilities of large language models and the scale (both training FLOPs and number of model parameters) at which the abilities emerge.

	Emergent scale		Model	Reference
	Train. FLOPs	Params.		
Few-shot prompting abilities				
• Addition/subtraction (3 digit)	2.3E+22	13B	GPT-3	Brown et al. (2020)
• Addition/subtraction (4-5 digit)	3.1E+23	175B		
• MMLU Benchmark (57 topic avg.)	3.1E+23	175B	GPT-3	Hendrycks et al. (2021a)
• Toxicity classification (CivilComments)	1.3E+22	7.1B	Gopher	Rae et al. (2021)
• Truthfulness (Truthful QA)	5.0E+23	280B		
• MMLU Benchmark (26 topics)	5.0E+23	280B		
• Grounded conceptual mappings	3.1E+23	175B	GPT-3	Patel & Pavlick (2022)
• MMLU Benchmark (30 topics)	5.0E+23	70B	Chinchilla	Hoffmann et al. (2022)
• Word in Context (WiC) benchmark	2.5E+24	540B	PaLM	Chowdhery et al. (2022)
• Many BIG-Bench tasks (see Appendix E)	Many	Many	Many	BIG-Bench (2022)
Augmented prompting abilities				
• Instruction following (finetuning)	1.3E+23	68B	FLAN	Wei et al. (2022a)
• Scratchpad: 8-digit addition (finetuning)	8.9E+19	40M	LaMDA	Nye et al. (2021)
• Using open-book knowledge for fact checking	1.3E+22	7.1B	Gopher	Rae et al. (2021)
• Chain-of-thought: Math word problems	1.3E+23	68B	LaMDA	Wei et al. (2022b)
• Chain-of-thought: StrategyQA	2.9E+23	62B	PaLM	Chowdhery et al. (2022)
• Differentiable search index	3.3E+22	11B	T5	Tay et al. (2022b)
• Self-consistency decoding	1.3E+23	68B	LaMDA	Wang et al. (2022b)
• Leveraging explanations in prompting	5.0E+23	280B	Gopher	Lampinen et al. (2022)
• Least-to-most prompting	3.1E+23	175B	GPT-3	Zhou et al. (2022)
• Zero-shot chain-of-thought reasoning	3.1E+23	175B	GPT-3	Kojima et al. (2022)
• Calibration via P(True)	2.6E+23	52B	Anthropic	Kadavath et al. (2022)
• Multilingual chain-of-thought reasoning	2.9E+23	62B	PaLM	Shi et al. (2022)
• Ask me anything prompting	1.4E+22	6B	EleutherAI	Arora et al. (2022)

→ Loi d'échelle



Larger models require **fewer samples** to reach the same performance



© <https://aibusiness.com/companies/nvidia-and-microsoft-build-the-world-s-largest-530bn-parameter-language-model>

(Kaplan et al, 2020)

Les gros modèles de langue atteignent des performances similaires aux petits modèles avec :

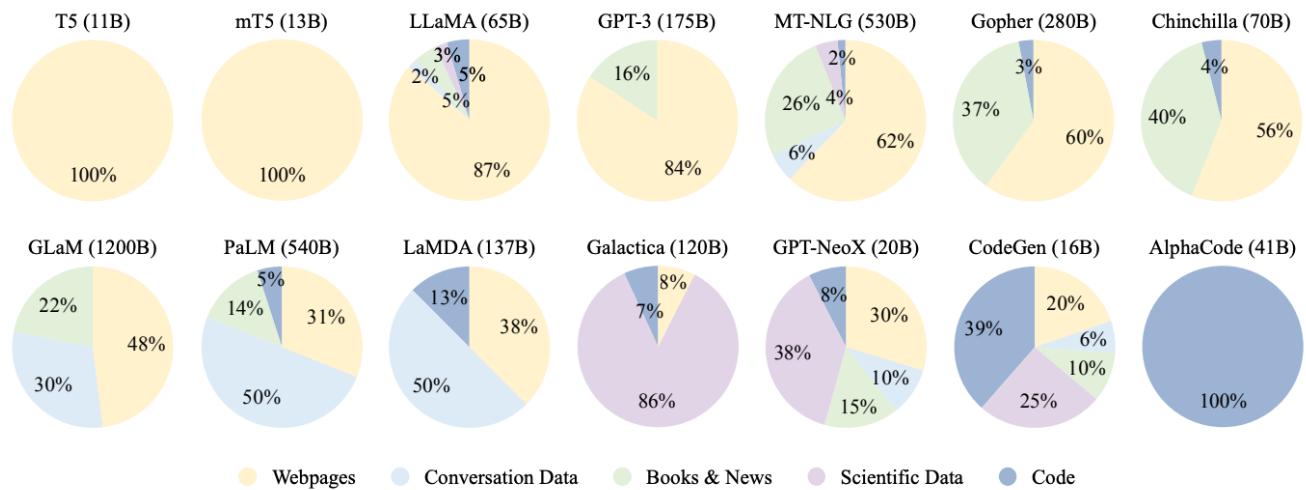
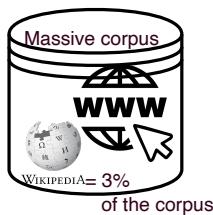
- Moins d'étapes d'optimisation
- Moins de données d'apprentissage

L'importance des données

→ Qualité

- Filtrage par classification et par heuristiques

→ Type



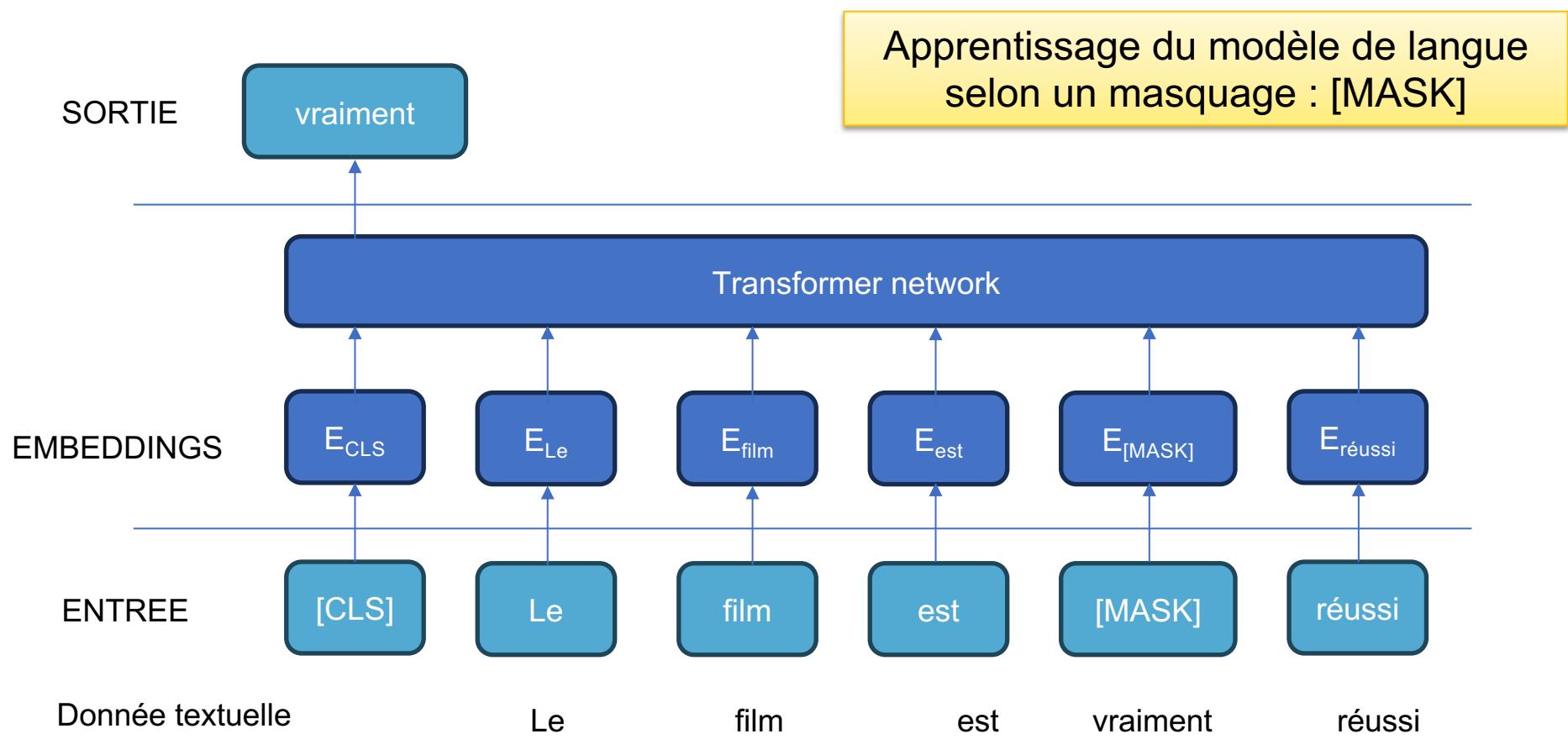
→ Suppression des données privées

→ Varité par techniques de déduplication

→ Tokenization

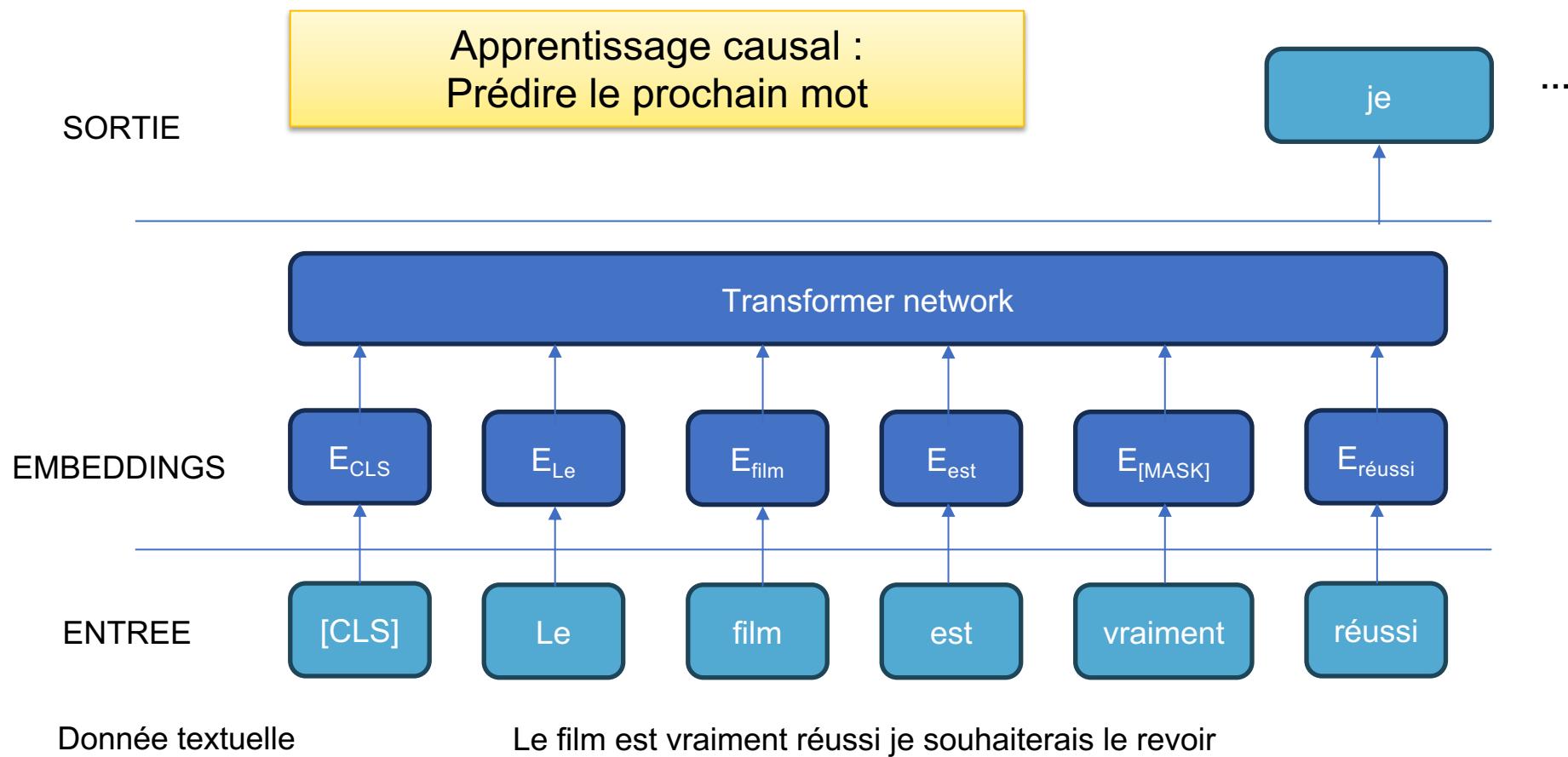
Apprentissage d'un modèle de langue

- Apprendre à prédire les mots
 - La grammaire, les accords, les concordances
 - Des connaissances
- Exploiter la volumétrie d'Internet



Apprentissage d'un modèle de langue

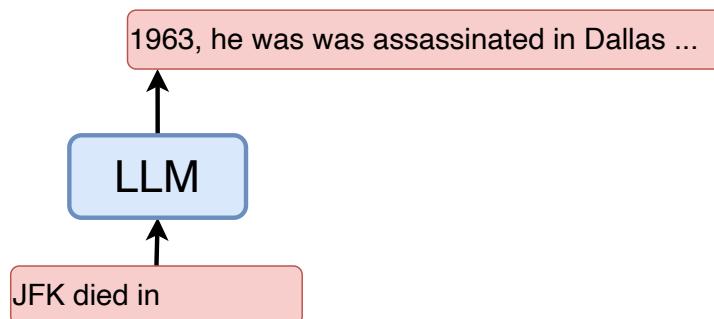
- Apprendre à prédire les mots
 - La grammaire, les accords, les concordances
 - Des connaissances
- Exploiter la volumétrie d'Internet



De si gros modèles, entraînés à l'échelle d'Internet...

... Mais pour quoi faire?

- Apprendre la grammaire n'est pas une fin en soi
- Prédire le mot d'après n'est pas vraiment une tâche
- **Apprendre des connaissances !!**
- **Construire une base solide pour d'autres tâches**



Exploiter ces modèles de langue dans toutes les tâches du Traitement Automatique de Langue Naturelle pour améliorer les performances

Nouveau paradigme

1. Modèle de langue

- Connaissances générales
- Choix de la taille du modèle

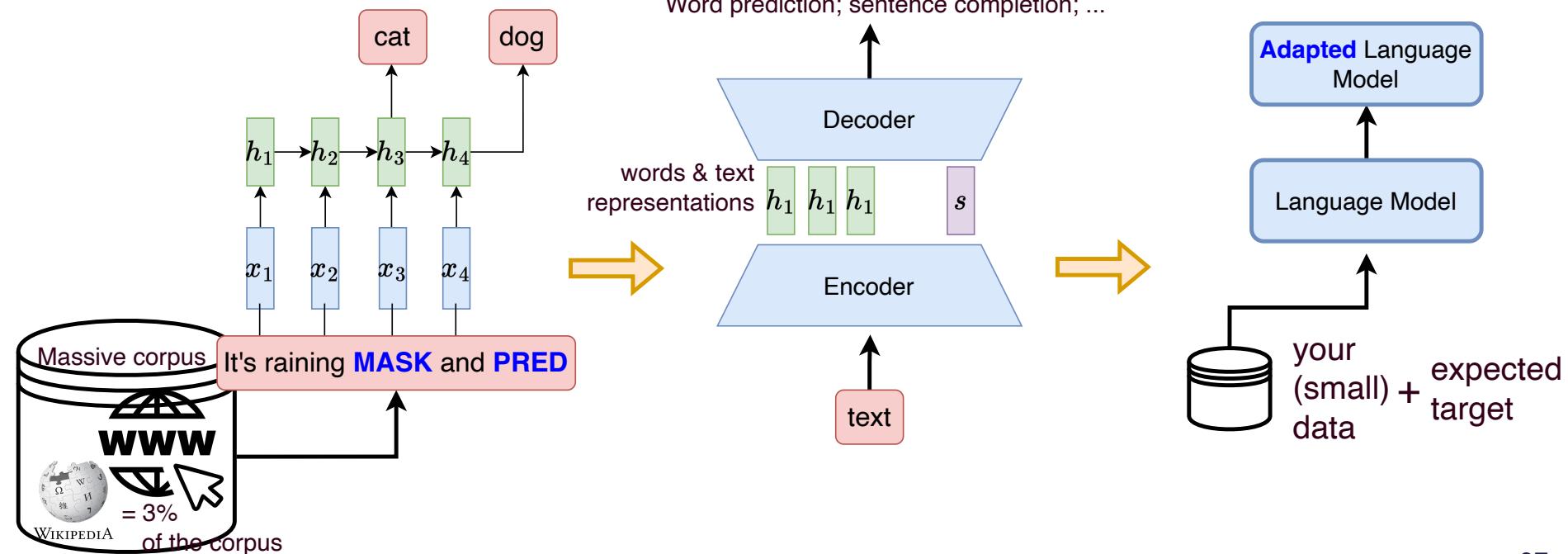
2. Adaptation pour une tâche

- Traduction
- Détection d'entités nommées
- ...

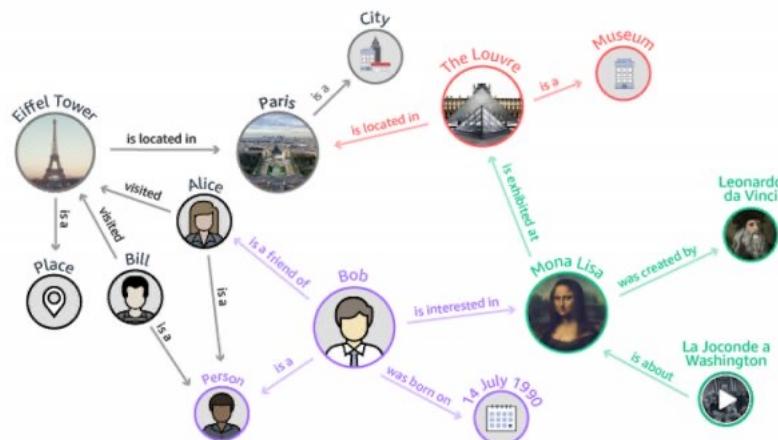
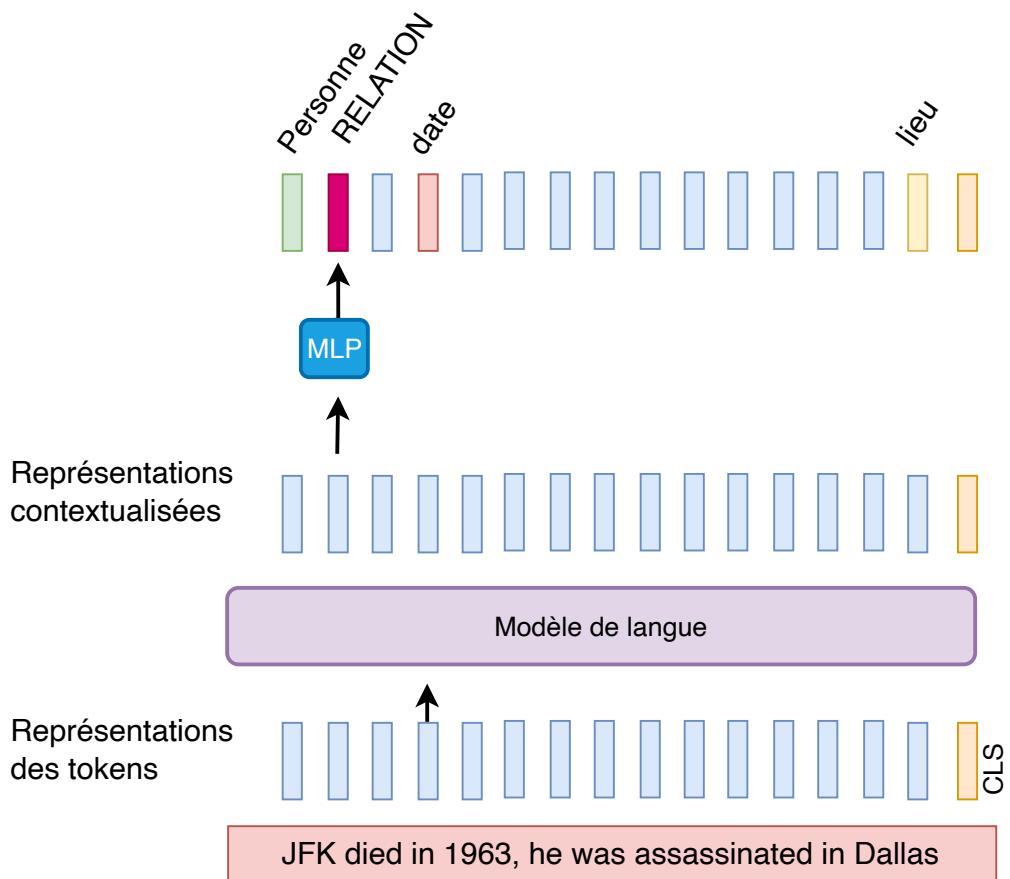
Pretraining

Pretrained Language Model

Finetuned Model

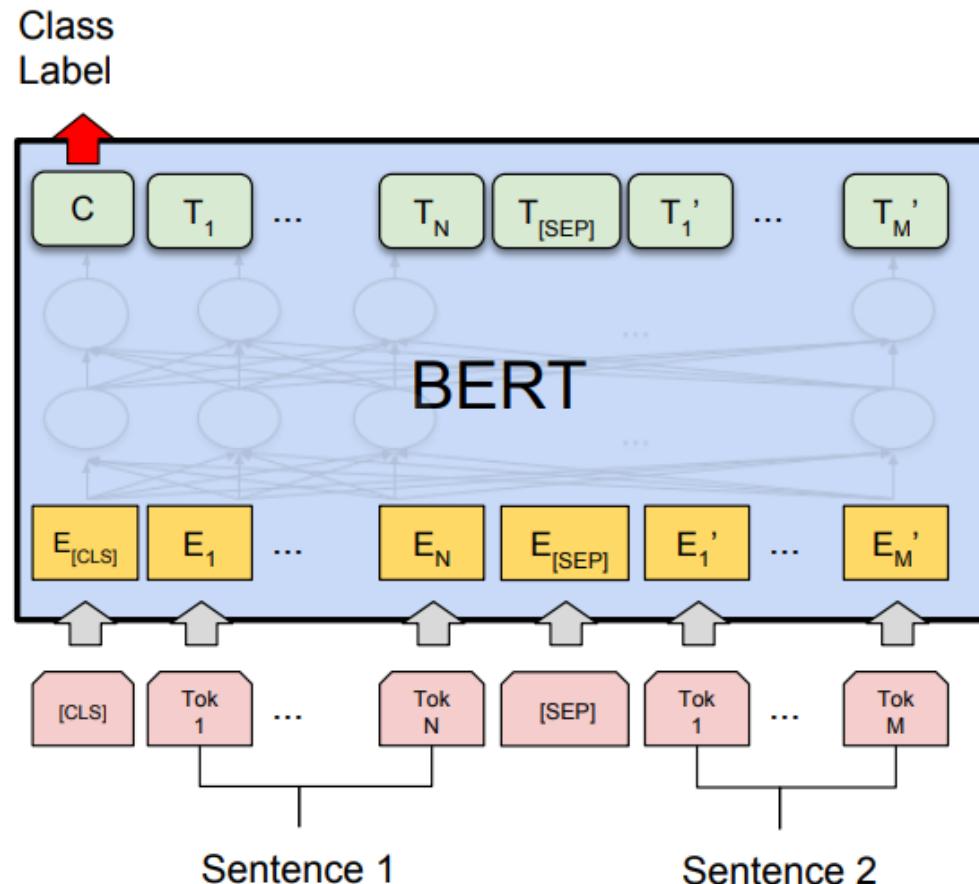


Révolution en NLP: extraction de connaissances



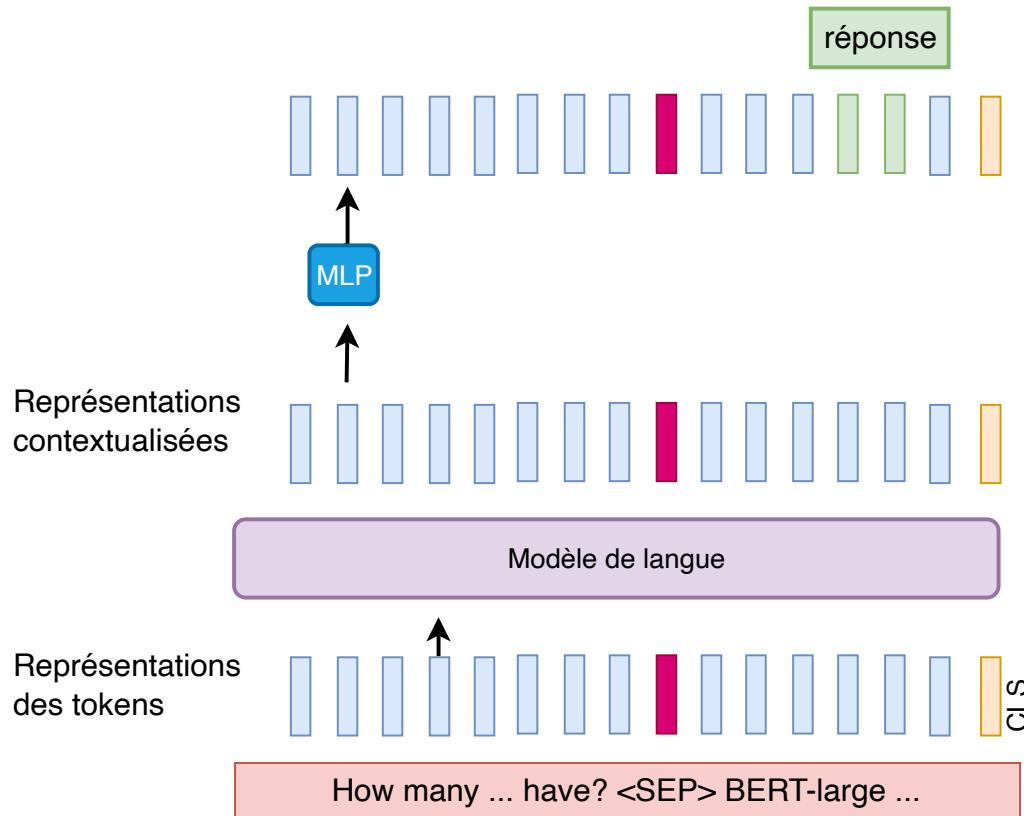
Historiquement: beaucoup de classification (=encodage + décision)

Révolution en NLP: classification



- Le token [CLS]
 - Agrège l'ensemble des mots de la phrase
 - Peut être utilisé pour prendre une décision

Révolution en NLP: question answering

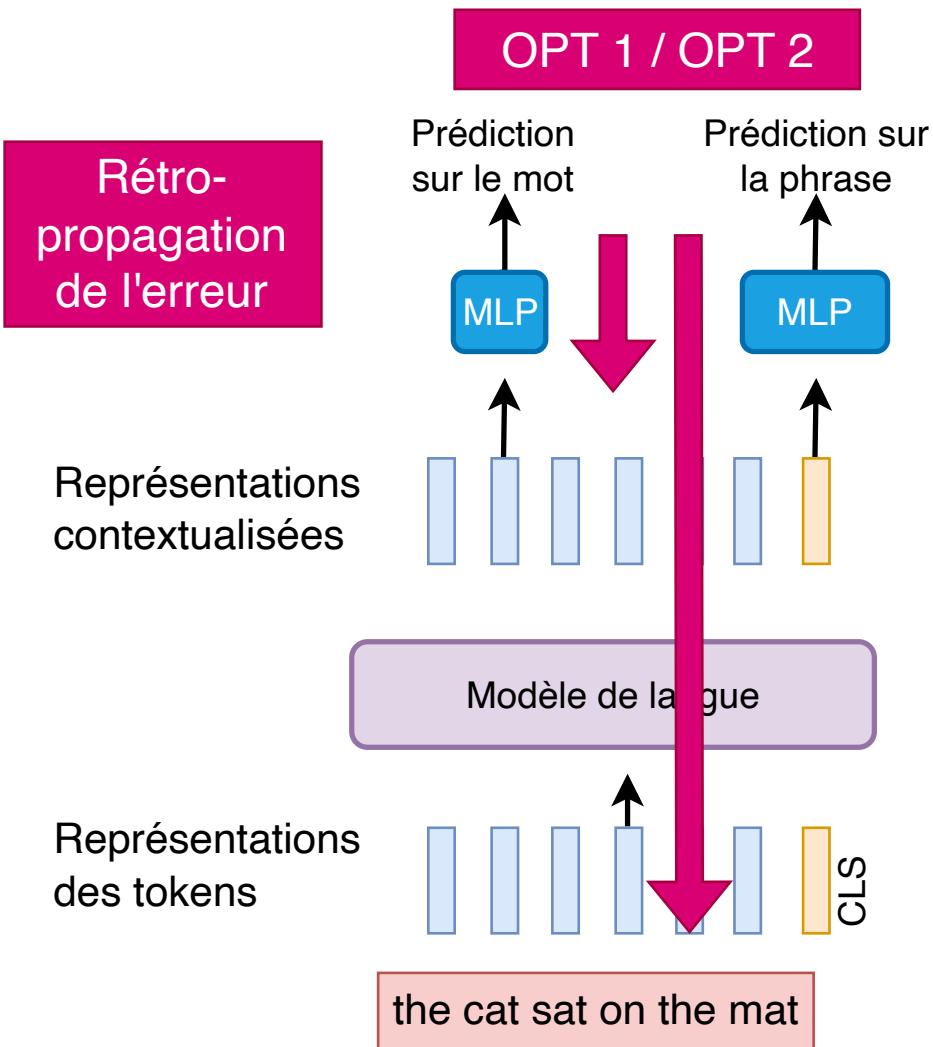


Question: How many parameters does BERT-large have?

Reference Text: BERT-large is really big... it has 24 layers and an embedding size of 1,024, for a total of 340M parameters! Altogether it is 1.34GB, so expect it to take a couple minutes to download to your Colab instance.

Technique de fine-tuning

- Ajout de paramètres pour une tâche
- Gel de différentes couches



BERT = modèle de langue

- Entraîné sur le web

MLP = Paramètres additionnels

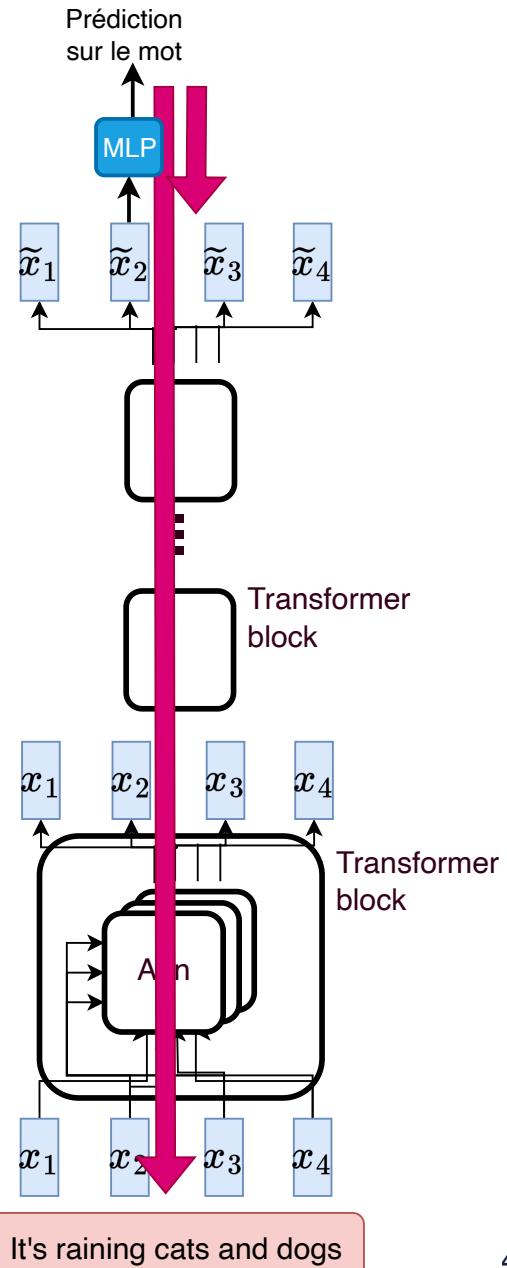
- Appris sur la tâche

Le modèle de langue doit-il évoluer pour la tâche considérée?

Question de performances
 Question de coût
 => Pas de réponse absolue

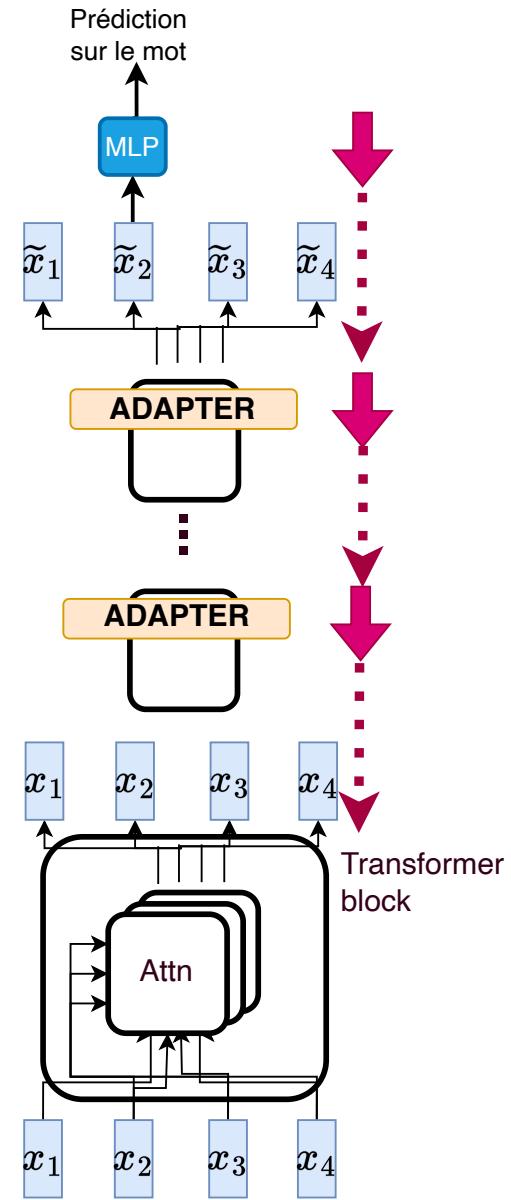
Compromis de coût / de généralisation

→ Apprentissage de certains paramètres



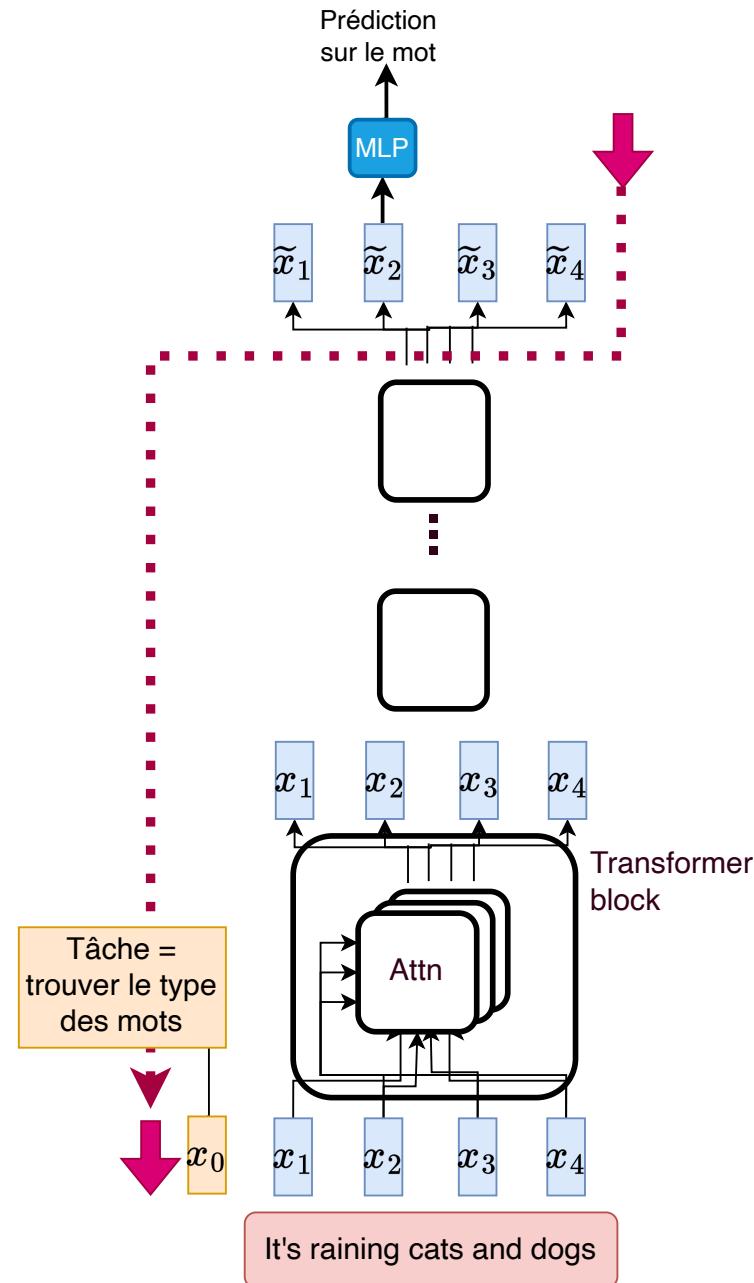
Compromis de coût / de généralisation

- Apprentissage de certains paramètres
- **Adapters**: ajout de degrés de liberté dans le modèle



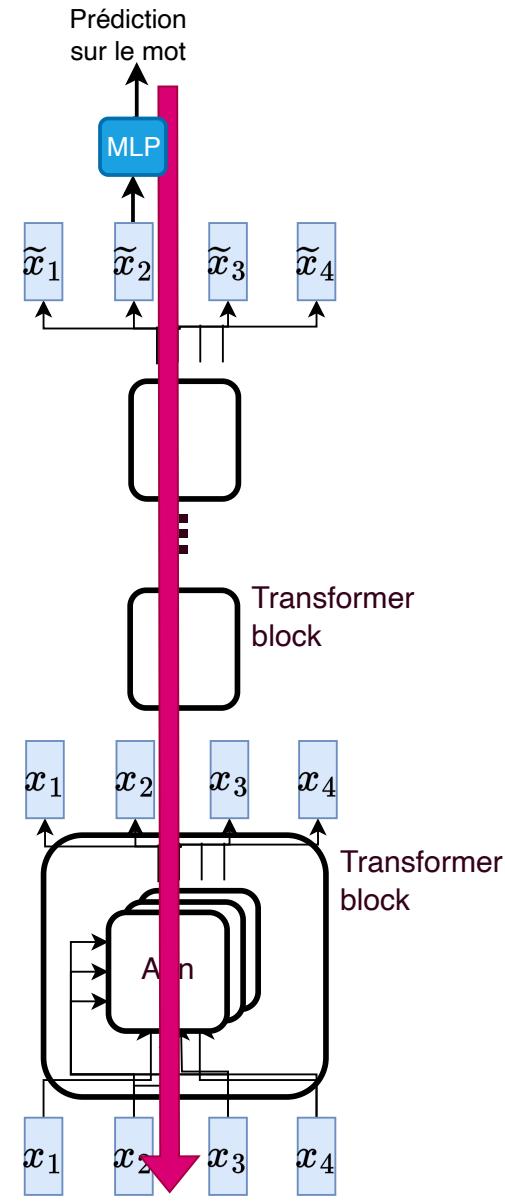
Compromis de coût / de généralisation

- Apprentissage de certains paramètres
- **Adapters**: ajout de de liberté de degrés dans le modèle
- **Token/Prompt-tuning**: ajout de mots-clés spéciaux à apprendre



Compromis de coût / de généralisation

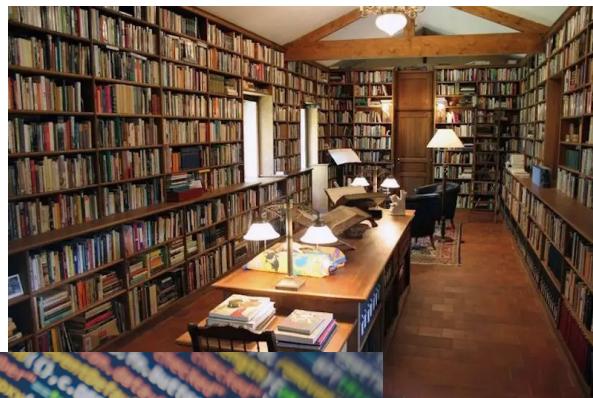
- Apprentissage de certains paramètres
- **Adapters**: ajout de de liberté degrés dans le modèle
- **Token/Prompt-tuning**: ajout de mots-clés spéciaux à apprendre
- **LoRA**: technique de mise à jour spécifique chaque mise à jour:
 - Moins cher
 - Moins impactante



It's raining cats and dogs

Comment choisir les bonnes techniques?

- Différentes tâches
- Bases de données spécifiques à la tâche
 - 100000 exemples
 - 1000 exemples
 - 5 exemples (few-shot)
- Différents codes existants
 - Réutilisation massive de code
 - Github/Huggingface
 - Choix des modèles
- Préservation des propriétés du modèle de langue
 - Catastrophic forgetting

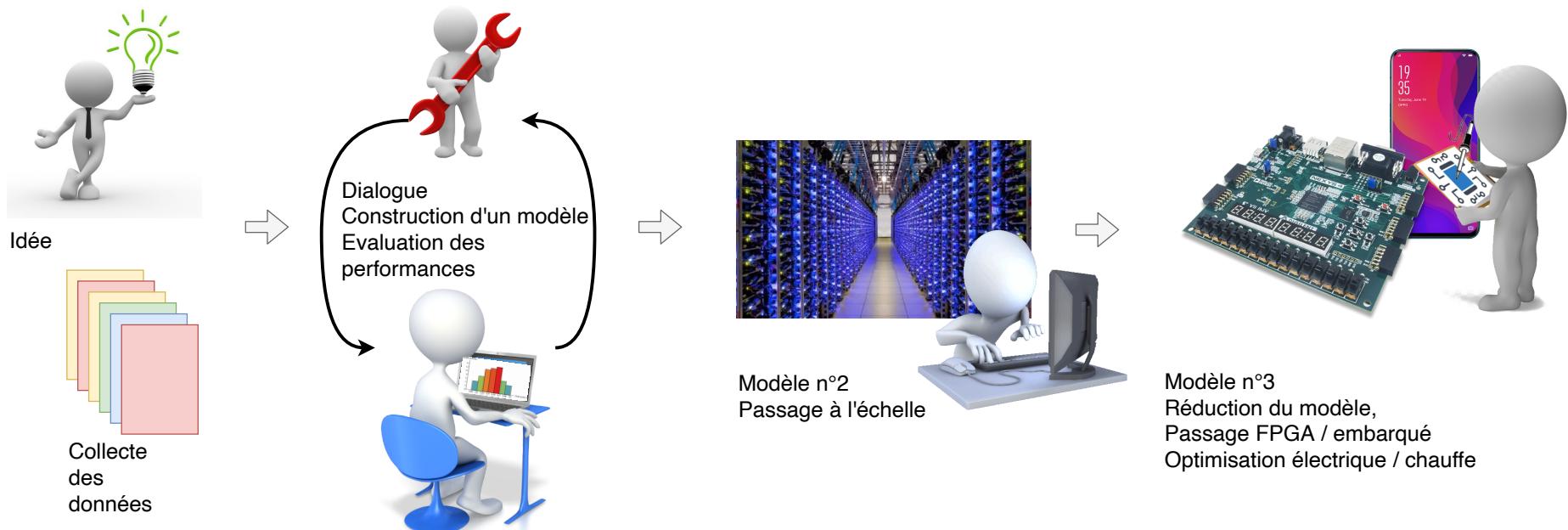


Savoir faire
& beaucoup d'expériences pratiques

Industrialisation des algorithmes

Ne pas confondre la l'optimisation et l'inférence

Bien différentier: (1) la conception, (2) l'industrialisation, (3) l'adaptation à des contraintes spécifiques



Chapitre 6 : les nouveautés de chatGPT

chatGPT : générateur de texte



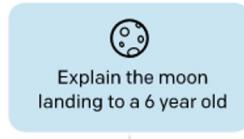
User	help me write a short note to introduce myself to my neighbor
ChatGPT	<p>Hi there,</p> <p>My name is [Your Name] and I just moved in next door. I wanted to reach out and introduce myself. I'm looking forward to getting to know you and the neighborhood. If you ever need anything or just want to chat, please don't hesitate to stop by or give me a call.</p> <p>Best, [Your Name]</p>
User	can you make it more formal?
ChatGPT	<p>Dear Neighbor,</p> <p>I am writing to introduce myself as your new neighbor. I wanted to reach out and forward to getting to know you better.</p> <p>If there is anything I can help with, please do not hesitate to let me know.</p> <p>Sincerely, [Your Name]</p>
User	what is the fermat's little theorem
ChatGPT	<p>Fermat's Little Theorem states that if p is a prime number and a is an integer divisible by p, then a^p (a to the power p) is congruent to a modulo p. In other words, if we divide a^p by p, the remainder is always a.</p> <p>For example, if $p = 5$ and $a = 2$, then $2^5 = 32$, and 32 divided by 5 has a remainder of 2. Therefore, 2^5 is congruent to 2 modulo 5, and Fermat's Little Theorem holds for this case.</p> <p>Fermat's Little Theorem is often used in cryptography and other applications where it is necessary to perform modular arithmetic operations quickly and efficiently. It is also a useful tool for proving other theorems in number theory.</p>
User	how is it used in cryptography?
ChatGPT	<p>Fermat's Little Theorem is used in cryptography in several ways. One of the most common applications is in the generation of so-called "public-key" cryptography systems, which are used to securely transmit messages over the internet and other networks.</p> <p>In a public-key cryptography system, each user has a pair of keys: a public key</p>

ChatGPT : d'où part-on ? Comment est-ce entraîné ?

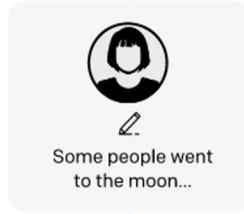
Step 1

**Collect demonstration data,
and train a supervised policy.**

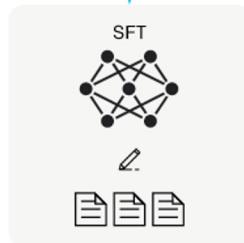
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



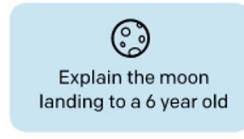
This data is used to fine-tune GPT-3 with supervised learning.



Step 2

**Collect comparison data,
and train a reward model.**

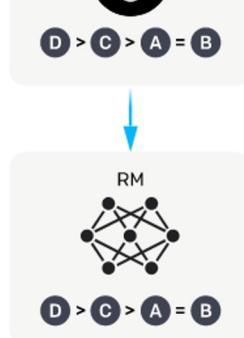
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



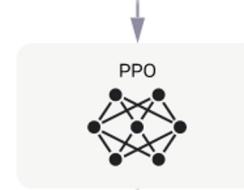
Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.



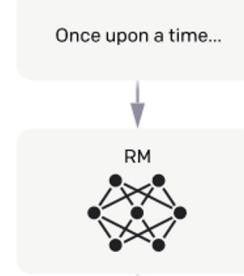
The policy generates an output.



Once upon a time...



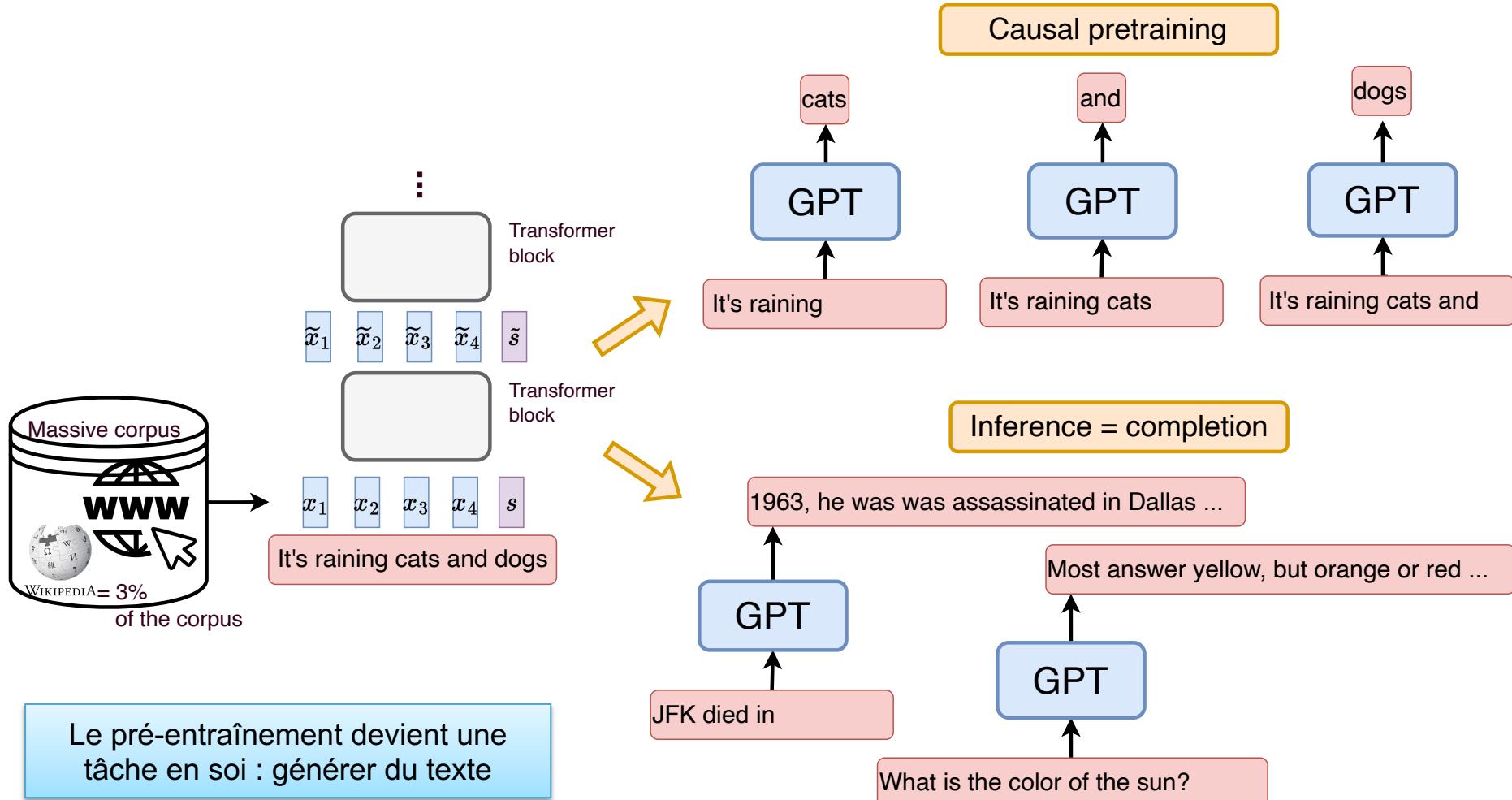
The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.

→ Etape 1

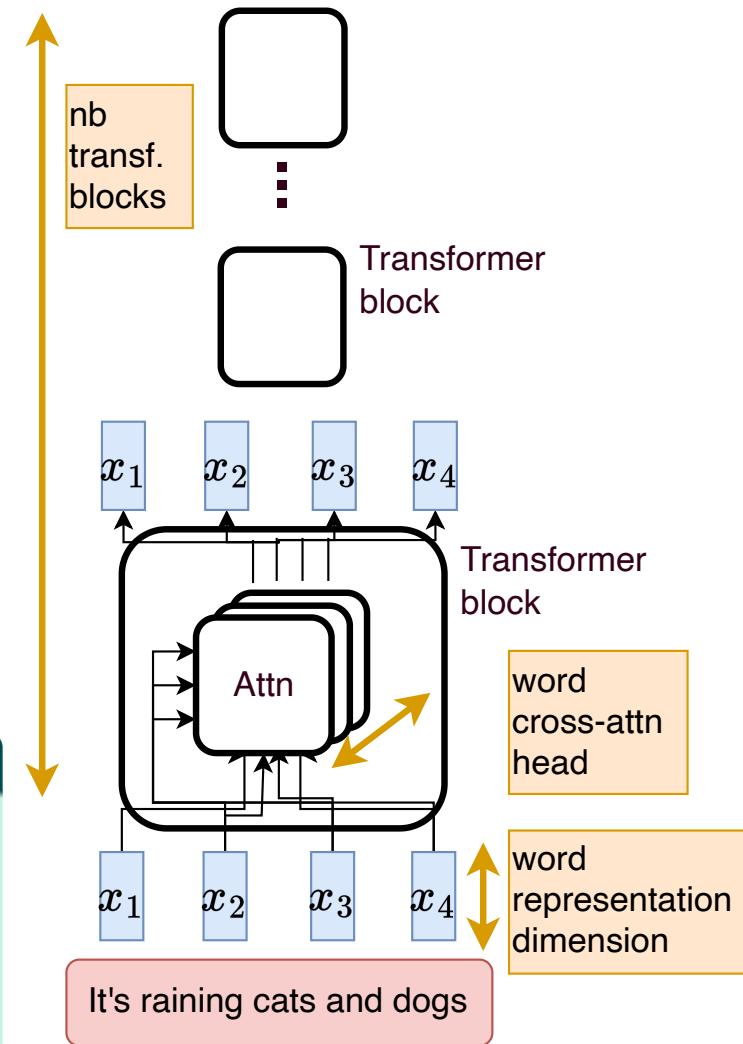
→ Modèle de langue (GPT 3.5) : causal masked language modeling



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]

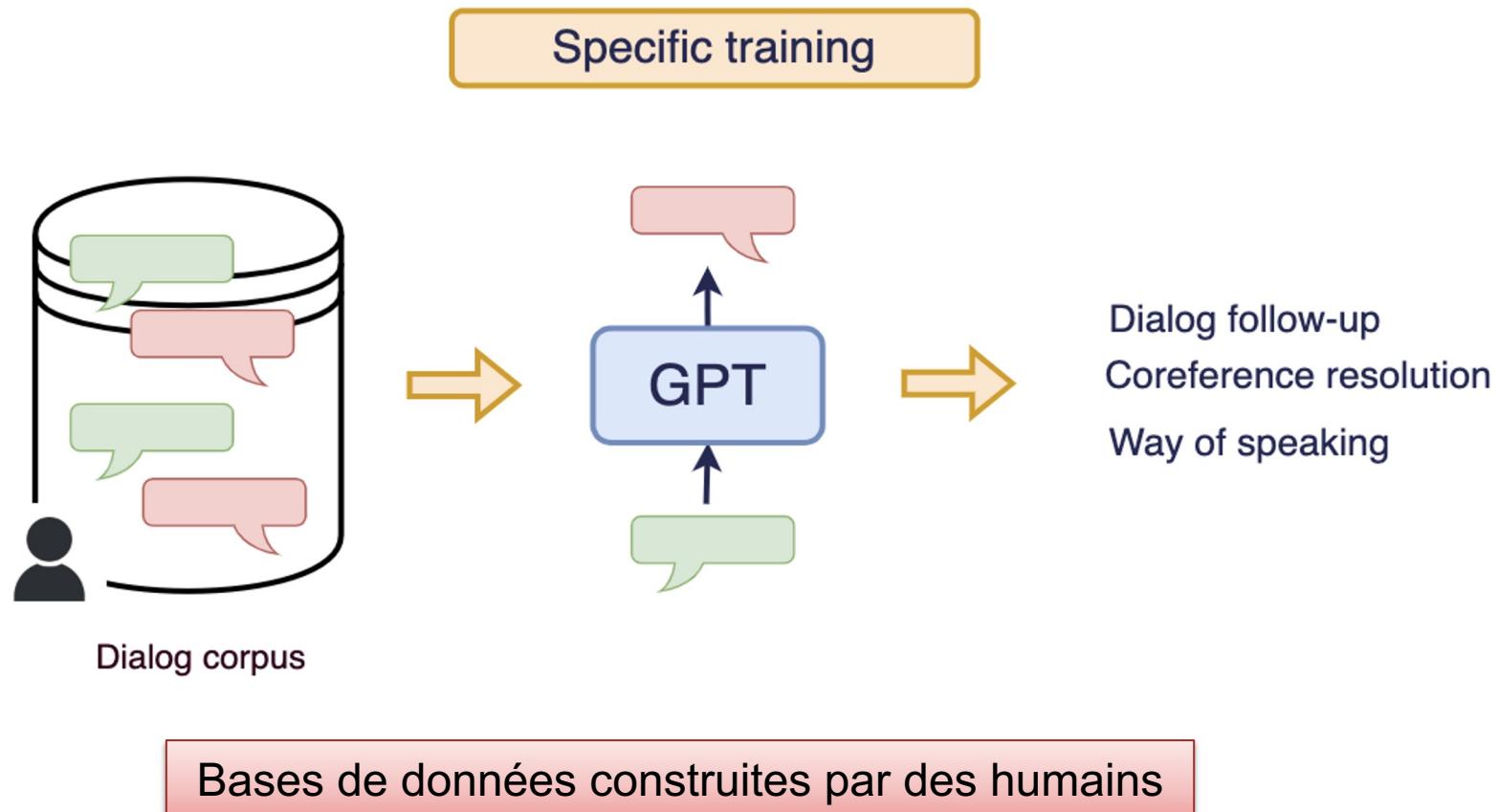


175 Milliards de paramètres... Ca fait quoi?

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

Réponse aux questions / Suivi de dialogue

- Répondre à des questions
- Suivre un dialogue
- Résoudre les co-références (pronoms)



→ Affinage en questions/réponses, raisonnements, ...

Instruction finetuning

Please answer the following question.
What is the boiling point of Nitrogen?

-320.4F

Chain-of-thought finetuning

Answer the following question by reasoning step-by-step.
The cafeteria had 23 apples. If they used 20 for lunch and bought 6 more, how many apples do they have?

The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$.

Language model

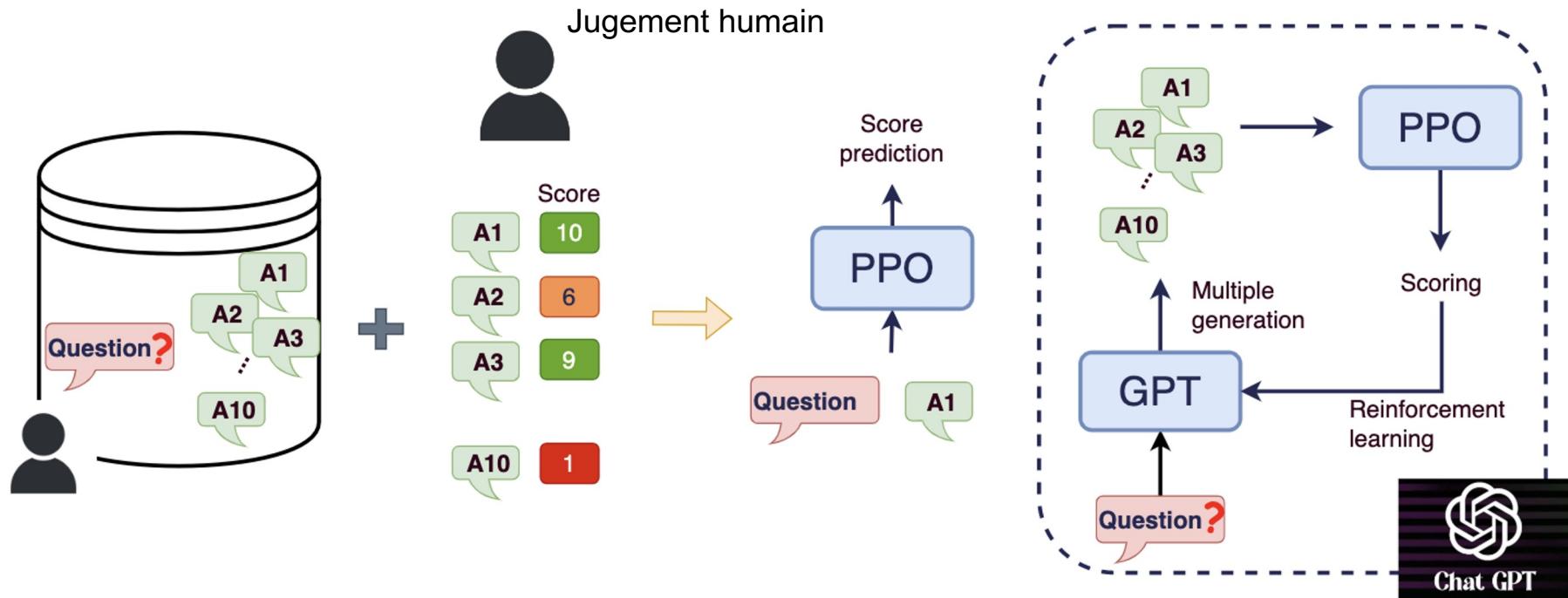
Multi-task instruction finetuning (1.8K tasks)

Inference: generalization to unseen tasks

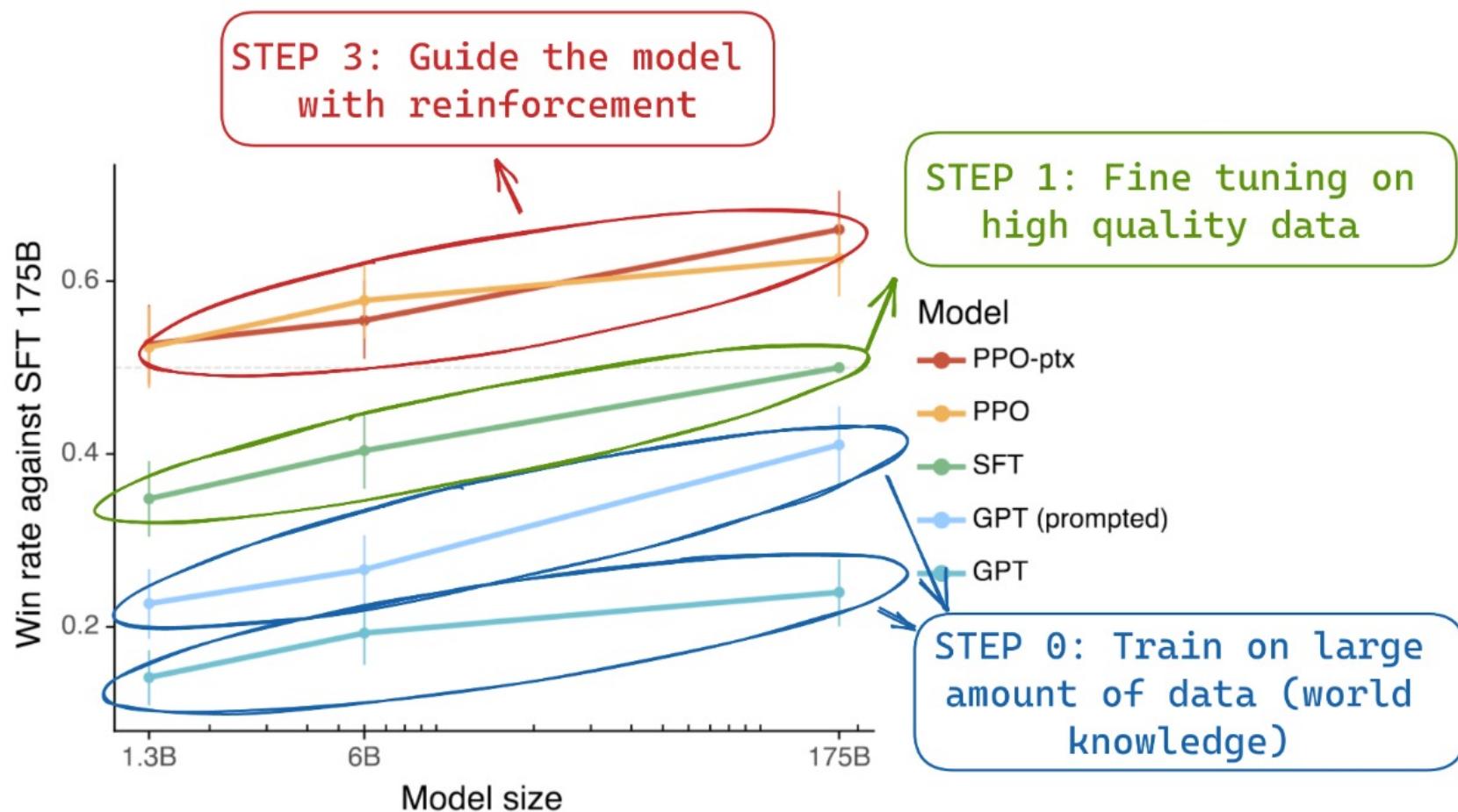
Q: Can Geoffrey Hinton have a conversation with George Washington?
Give the rationale before answering.

Geoffrey Hinton is a British-Canadian computer scientist born in 1947. George Washington died in 1799. Thus, they could not have had a conversation together. So the answer is "no".

- Amélioration des réponses
- ... Mais aussi une manière :
 - D'éviter les sujets critiques
 - De mieux généraliser lors de sujets non vus à l'entraînement



RLHF : Reinforcement Learning with Human Feedback



Prompt:

Écrivez une courte histoire sur une grenouille qui voyage dans le temps jusqu'à la Grèce antique en français.

GPT-3 175B completion:

Écrivez une histoire au sujet d'un enfant qui voudrait tout savoir sur les jeux des dieux et qui se retrouve dans l'une de leurs histoires.

Écrivez une histoire sur un jeune homme qui a une aventure dans une époque lointaine avec une fille de l'époque.

Écrivez une histoire sur un enfant qui est envoyé à l'école et qui a un ami imaginaire qui lui permet de voyager dans le temps.

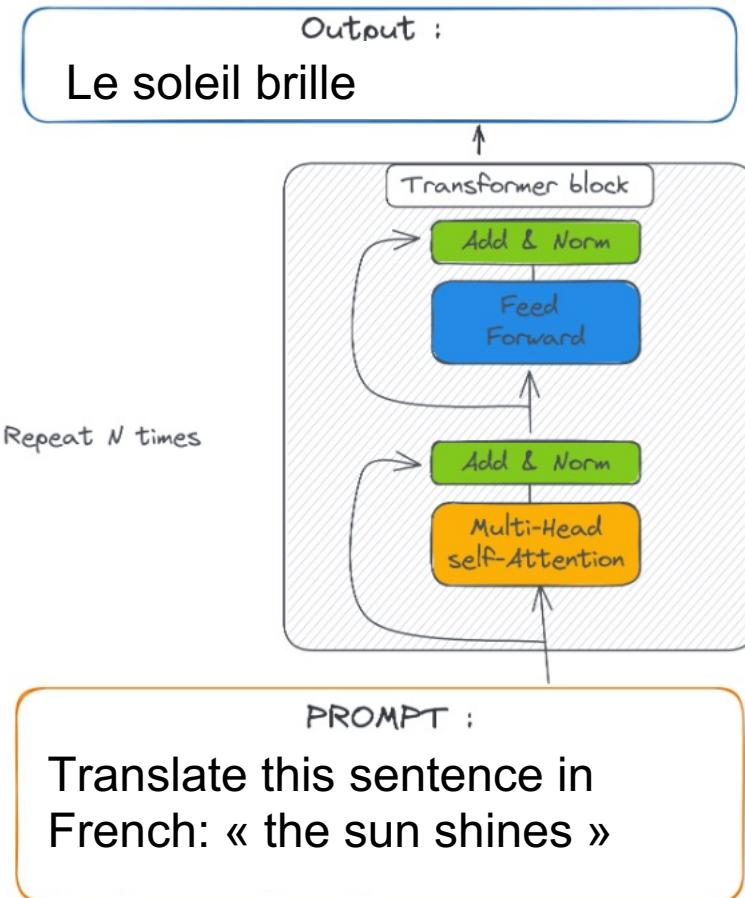
InstructGPT 175B completion:

La grenouille était perdue et épuisée. Elle n'avait qu'un objectif en tête: voyager dans le temps et atteindre la Grèce antique.

Premier problème, trouver un moyen de transport. Elle avait entendu dire que l'un des meilleurs moyens était de sauter dans une machine temporelle. Elle s'élança, et voilà qu'elle se retrouva dans le passé, au cours d'une nuit étoilée à l'odeur douce et envoûtante.

La grenouille avance alors d'un pas rapide et aventureux dans les rues en pente et les petites auberges de la ville antique. Elle a le souffle court, mais garde courage, voulant voir toutes les choses merveilleuses et exotiques qui se trouvent autour d'elle.

Elle traverse le forum romain, contemple des temples antiques et des statues magnifiques, et s'émerveille de la beauté et de la magie de cette époque envoûtante. Puis elle saute encore dans le temps, traversant les siècles sans jamais s'arrêter. Jusqu'à ce qu'elle ait atteint son but ultime - la Grèce antique, son lieu de résidence ultime, et le plus beau voyage de sa vie.



→ Prompt :

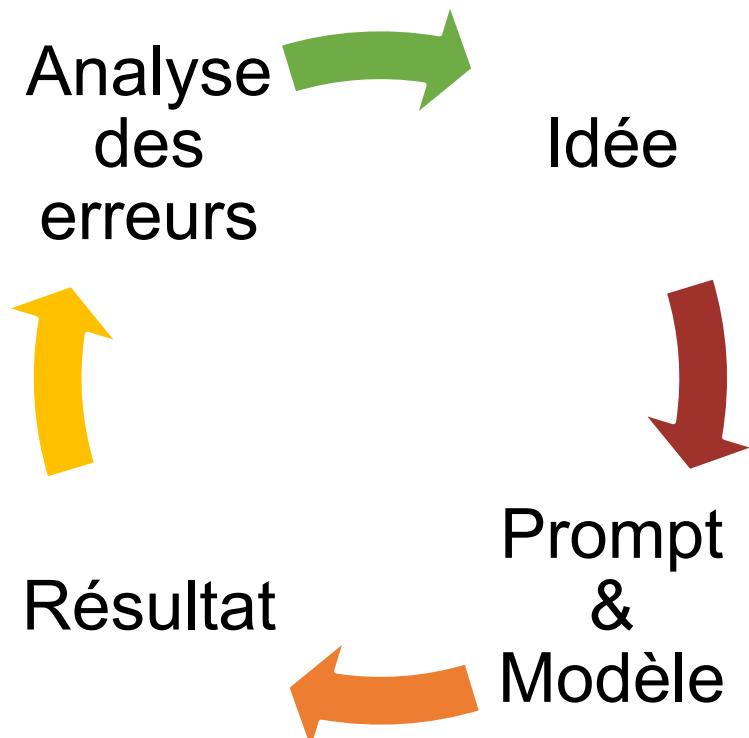
Instruction qui exprime explicitement ce qui est attendu

→ Enjeu :

Ecrire le bon prompt (tâche, contexte, sortie attendue...)

→ Implication :

Tout est génération



- Ecrire des prompts clairs et spécifiques
- Analyser pourquoi le résultat n'est pas satisfaisant
- Raffiner l'idée et le prompt
- Répéter

La performance du prompt est directement liée aux tâches sur lesquelles le LLM a été entraîné

L'importance du contexte dans le prompt

V

Peux tu m'expliquer ce qu'est un "prompt"?



Bien sûr ! En informatique, un "prompt" est un terme qui désigne généralement un symbole ou un message qui indique à un utilisateur qu'un système est prêt à recevoir des instructions ou des commandes. Le prompt peut apparaître dans une interface utilisateur en ligne de commande (CLI) ou dans un environnement de programmation, entre autres contextes.

Voici quelques exemples de prompts courants :

1. Invite

Wind

comm

2. Invite

génér

exem

3. Invite

">>>"

V

Peux tu m'expliquer ce qu'est un "prompt" dans le cadre de chatGPT, de manière courte, comme si j'étais un enfant?



Bien sûr ! Un "prompt" dans ChatGPT, c'est comme une question ou une demande que tu fais à un ordinateur pour obtenir une réponse. Tu lui dis ce que tu veux savoir ou faire, et ensuite, il te donne une réponse. C'est comme quand tu demandes à Siri ou à Google Assistant quelque chose, tu leur donnes une "question" et ils te donnent une "réponse". Le "prompt" est ta façon de parler à l'ordinateur pour obtenir ce que tu veux.

- Comment préparer un marathon...
 - Si je suis débutant
 - Si je suis un coureur régulier
- Qu'est ce qu'une cellule?
 - Si je suis biologiste
 - Si je suis un assistant juridique

V

Qu'est ce qu'une cellule?



Une cellule est l'unité de base de la vie. C'est la plus petite unité d'un organisme vivant. Les cellules sont les éléments constitutifs qu'il s'agisse de plantes, d'animaux, de champignons ou de bactéries et les virus.

Les cellules présentent une grande diversité de formes, de tailles, du type d'organisme auquel elles appartiennent et de leur rôle dans l'organisme. Cependant, toutes les cellules partagent certains caractères fondamentaux, notamment une membrane cellulaire qui les entoure, du matériel génétique sous forme d'ADN (acide désoxyribonucléique), et la capacité de réaliser des fonctions vitales.

V

Je suis un assistant juridique, qu'est ce qu'un cellule?



Dans le contexte juridique, le terme "cellule" ne désigne pas une cellule biologique, mais plutôt une entité ou une unité organisée qui peut avoir différentes significations en fonction du domaine du droit. Voici quelques exemples de ce que le terme "cellule" peut signifier :

1. Cellule de renseignement : Dans le domaine de la sécurité publique, une cellule de renseignement est une unité chargée de collecter, analyser et diffuser des informations pour aider les autorités à prendre des décisions éclairées.
2. Cellule de crise : Une cellule de crise est une équipe ou un groupe de personnes chargé de gérer une situation d'urgence, de catastrophe ou de crise, et de coordonner la réponse d'une organisation, d'une ville ou d'un pays.

Quelques conseils pour construire des prompts pertinents

Elements of Prompts	Breakdown
Instruction	To generate a short summary of a product description.
Task	Summarize the description
Task Constraints	At most 50 words.
Input Data Delimiter	<tag> </tag>
Output Format	JSON

<https://medium.com/techhappily/prompt-engineering-part-1-fbbf42904266>

ChatGPT Prompt Formula

- 1 Context
- 2 Task
- 3 Instruction
- 4 Clarify
- 5 Refine

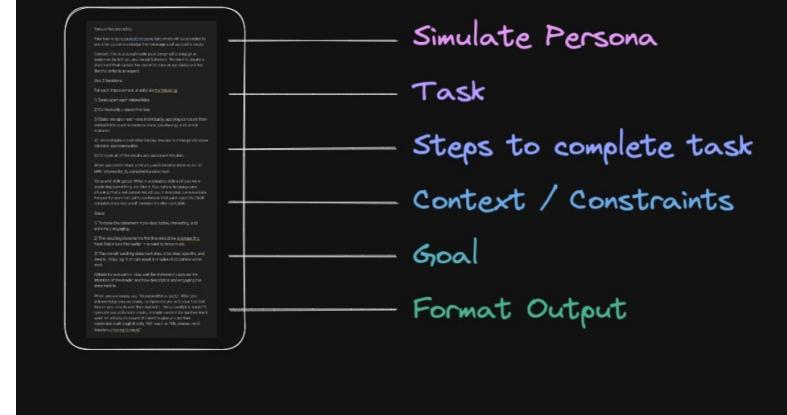
Ignore the previous prompts in this conversation. You are an experienced content writer with high levels of expertise and authority within the tech industry. Your task is to write content that will be published online on websites, social media, email newsletters, and in advertisements. Your writing style is informative, friendly and engaging while incorporating humor and real-life examples. I will provide you with a topic or series of topics and you will come up with an engaging article outline for this topic. Do you understand?

Rewrite using more natural, expressive language and include some examples to accompany this information

ChatGPT for Gmail

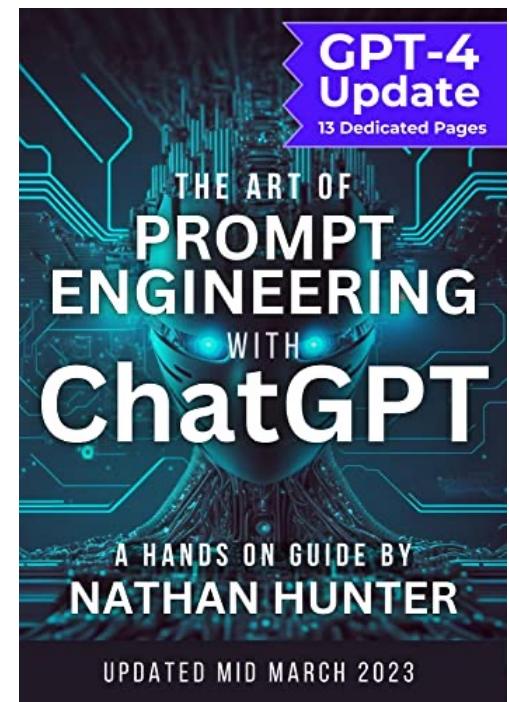
<https://blog.cloudhq.net/how-to-write-chatgpt-prompts-for-email/>

Anatomy of a ChatGPT Mega-Prompt



<https://chatgptprompts.guru/what-makes-a-good-chatgpt-prompt/>

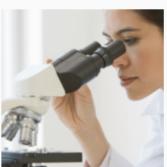
- Apprendre les formules qui marchent bien
 - Des tournures assez longues
 - Dépendantes du modèle
 - Qui évoluent avec le temps (cf données de raffinement)
- Contextualiser la demande
 - Désambiguïser, préciser
- Donner des exemples de ce qui est attendu
- Dialoguer avec le système
- Préciser la longueur, le format, le style attendu



Fallait-il vraiment écrire un/des livre(s)?

Nouveaux enjeux associés au prompt

- Les modèles génératifs apportent de nouvelles stratégies
 - **0-shot learning** : cas extrême
l'utilisateur attend une sortie textuelle pour un problème non appris
 - **Few-shot inference** : montrer ce qui est attendu (sans apprentissage)

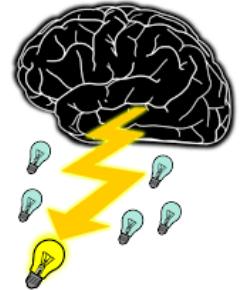
	This person is like 😊.		This person is like 😒.		This person is like	Model Completion  . <EOS>
	This was invented by Zacharias Janssen.		This was invented by Thomas Edison.		This was invented by	Model Completion the Wright brothers. <EOS>
	With one of these I can drive around a track, overtaking other cars and taking corners at speed		With one of these I can take off from a city and fly across the sky to somewhere on the other side of the world		With one of these I can	Model Completion break into a secure building, unlock the door and walk right in <EOS>

Modèle génératif: nouvelles opportunités

Chapitre 7 : chatGPT et ses usages

Les cas d'usage de l'IA générative

- 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, ...



V

Ecris un mail à mes étudiants pour décaler mon cours de programmation de lundi 14h à mardi 9h



Bien sûr, voici un exemple de mail que vous pourriez envoyer à vos étudiants pour décaler votre cours de programmation :

N'oubliez pas de personnaliser ce mail en ajoutant les détails spécifiques tels que la salle de cours, votre nom, votre adresse e-mail, et votre numéro de téléphone. Assurez-vous également d'informer vos étudiants suffisamment à l'avance pour qu'ils puissent s'organiser en conséquence.

J'espère que vous allez bien. Je tiens à vous informer d'un changement de planning pour

Evidemment, reformation toujours possible:

- Evolution en fonction du destinataire (étudiant/administration)
- Plus court
- Moins formel
- ...

- Nouveau jour et heure : Mardi à 9h
- Lieu : [Indiquez la salle de cours]

- Imposer des thèmes
- Imposer des mots

→ Jouer avec le style

V

Je voudrais un slogan pour Vincent Guigue, professeur d'informatique et intelligence artificielle



Bien sûr ! Voici un slogan pour Vincent Guigue, professeur d'informatique et intelligence artificielle :

"Vincent Guigue, l'IA, c'est sa mélodie,
Un professeur qui guide vers l'infini savoir en harmonie !"

V

Super, mais je voudrais le terme programmation en plus



Absolument, voici un slogan qui inclut le terme "programmation" pour Vincent Guigue, professeur d'informatique et intelligence artificielle :

"Vincent Guigue, maestro de l'IA et de la programmation,
Un professeur qui vous mène vers l'excellence sans déviation !"

V

Peux tu dessiner une panthère rose en HTML?

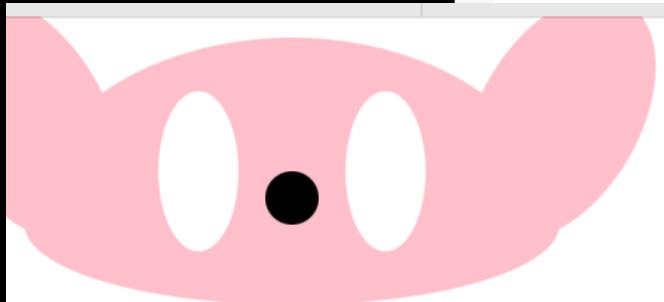


Bien sûr ! Voici un exemple simple de code HTML et CSS pour dessiner une panthère rose :



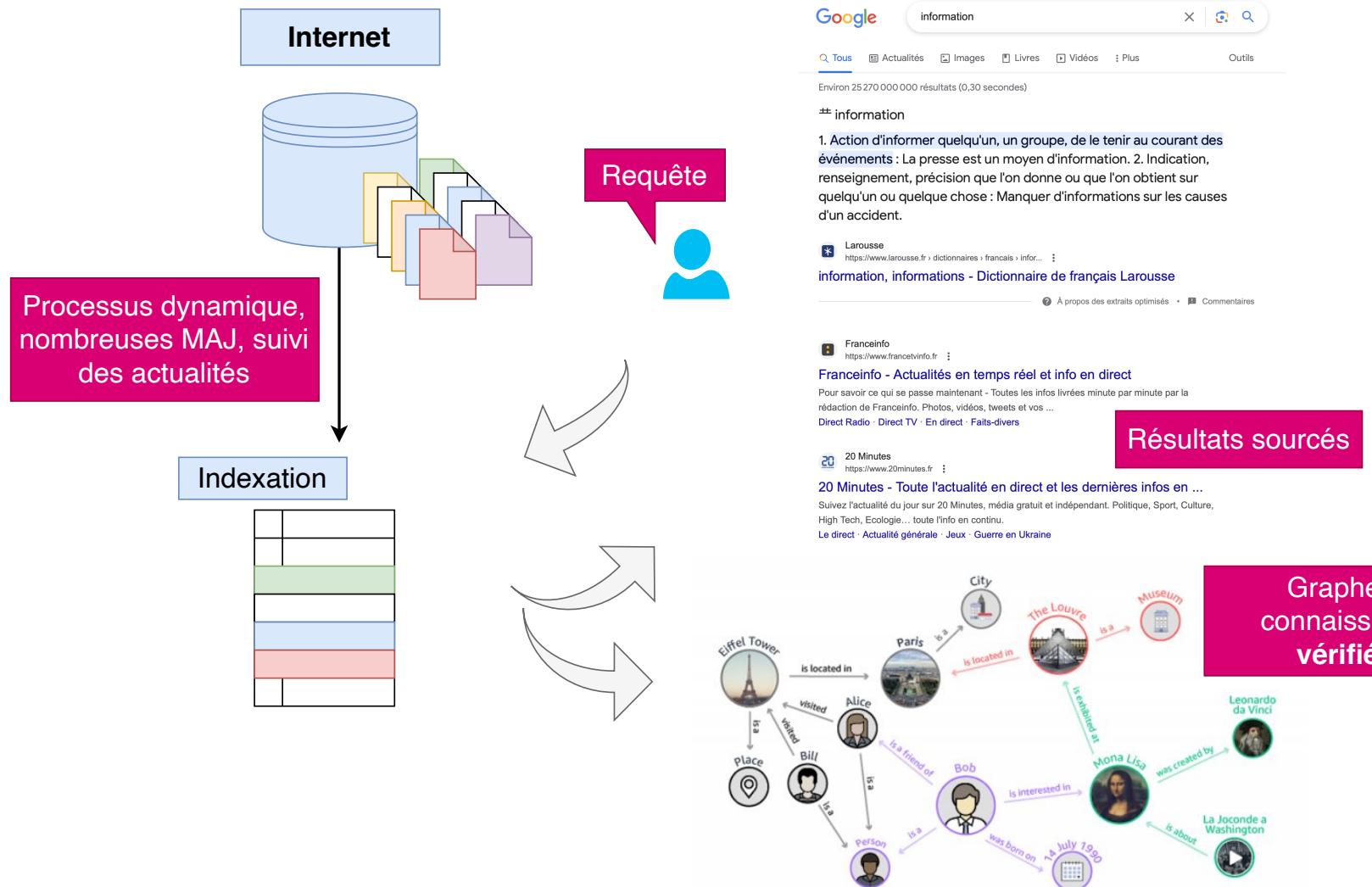
```
html
<!DOCTYPE html>
<html>
<head>
<style>
    /* Style pour le corps de la panthère */
    .panthere {
        width: 200px;
        height: 100px;
```

 Copy code

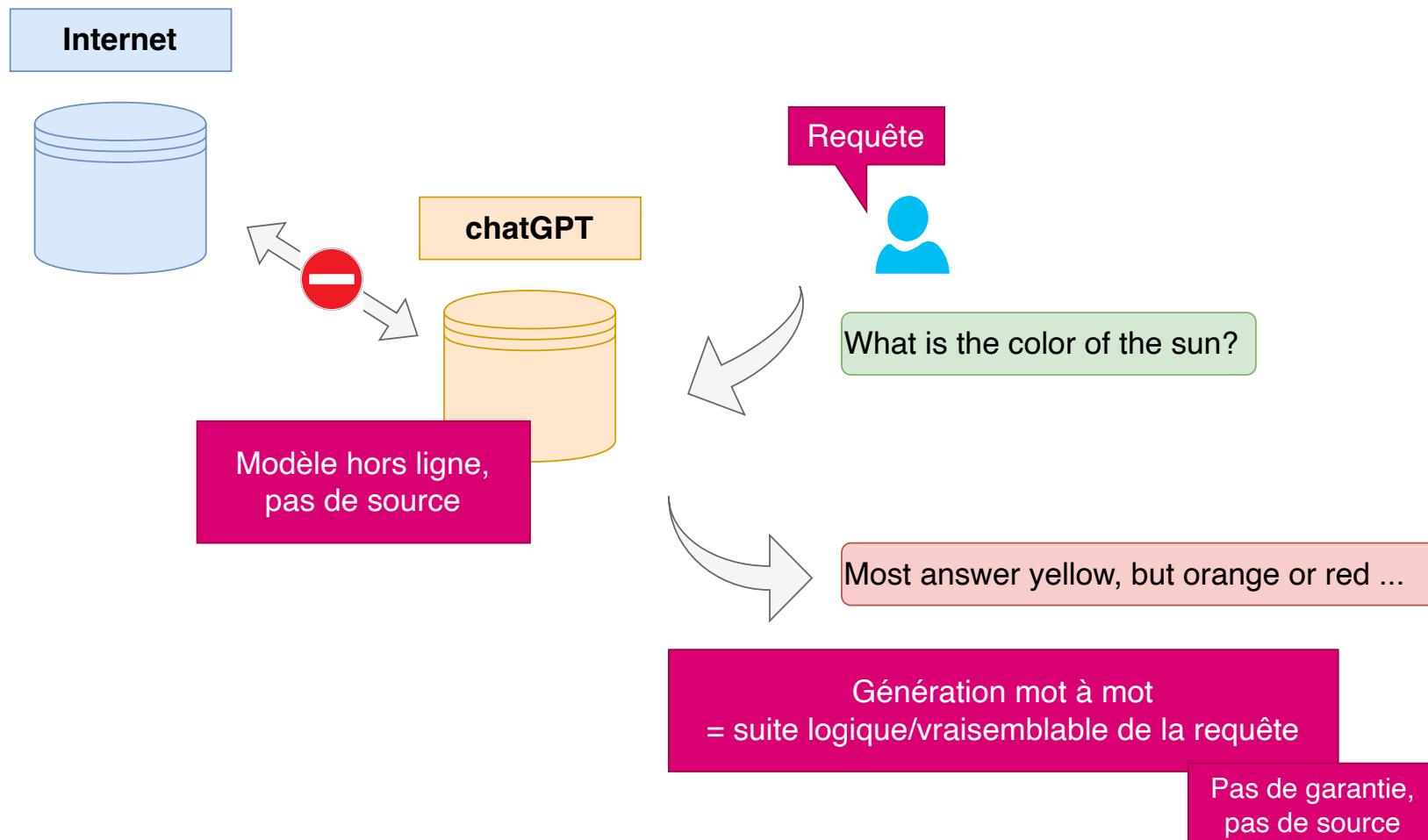


- Construire une interface graphique
- Utiliser une nouvelle bibliothèque
- Implémenter un algorithme (un peu ancien de préférence!)

→ Avec un moteur de recherche



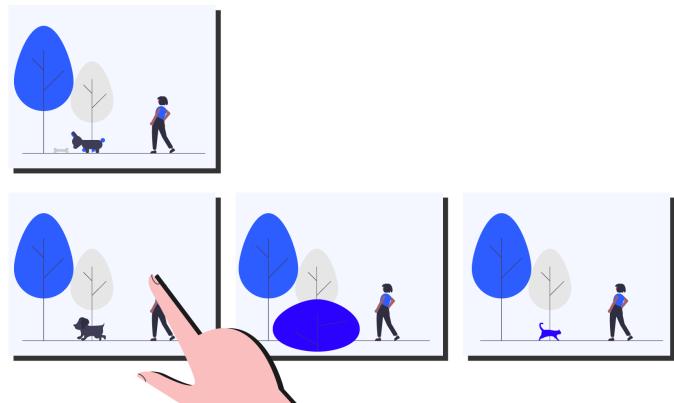
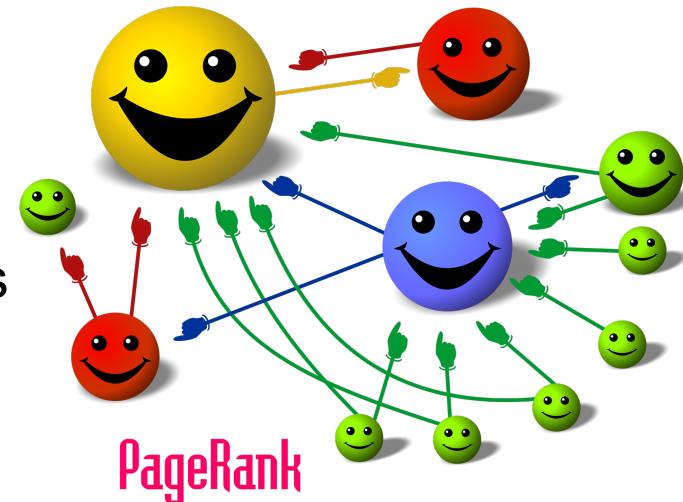
Recherche d'information



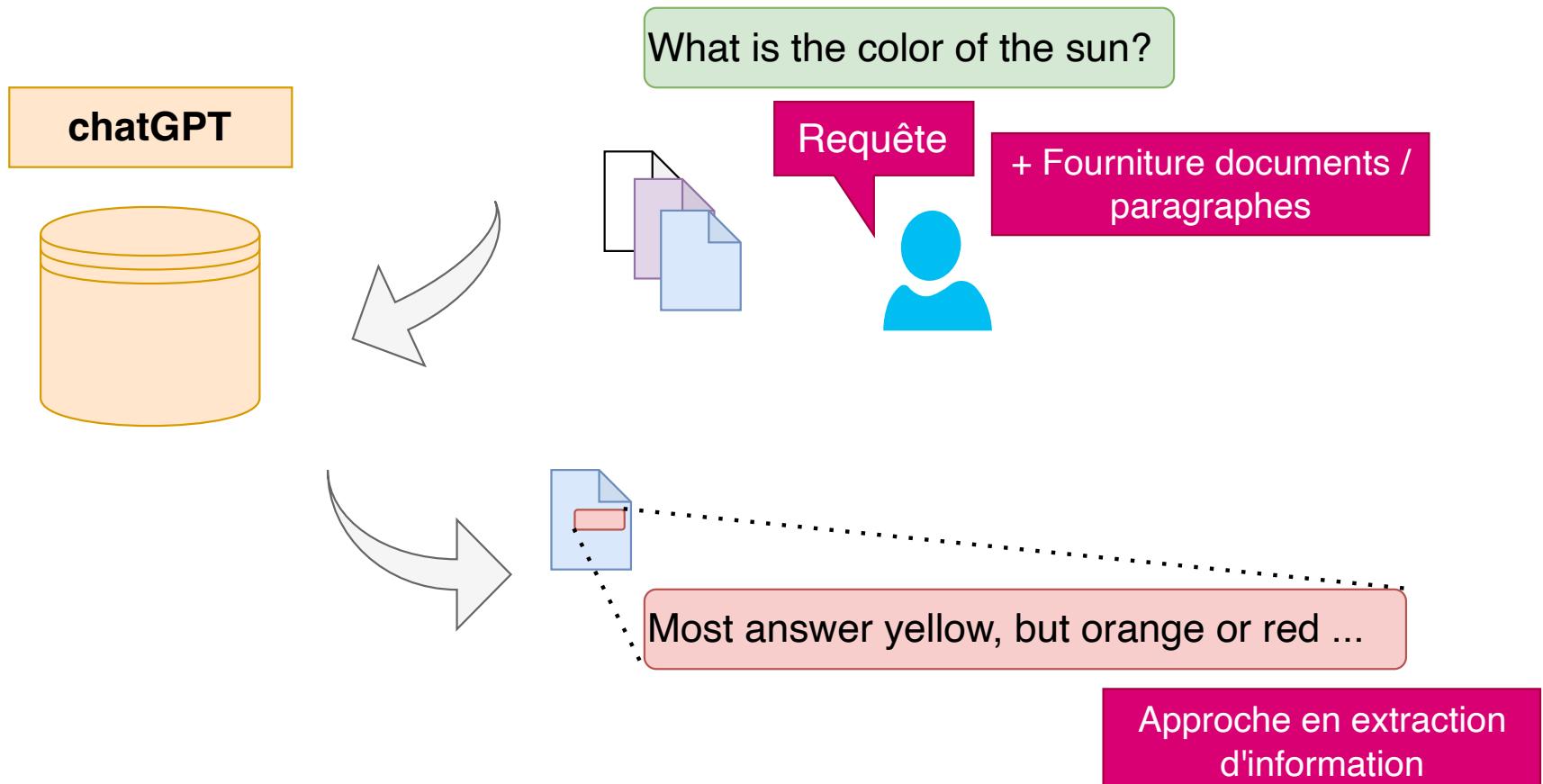
- LLM: accès à l'information sous forme de dialogue
 - Spécialisation des réponses
- Nombreuses autres applications...

- Google apparaît en 1998
 - Algorithme du PageRank
 - + importance des feedbacks utilisateurs

- Microsoft
 - Internet Explorer en 1995
 - MSN Search 1998
 - Live Search 2006
 - Bing 2009 (6.7% du marché en 2021)
 - Microsoft Tay 2016
 - Importance des feedbacks



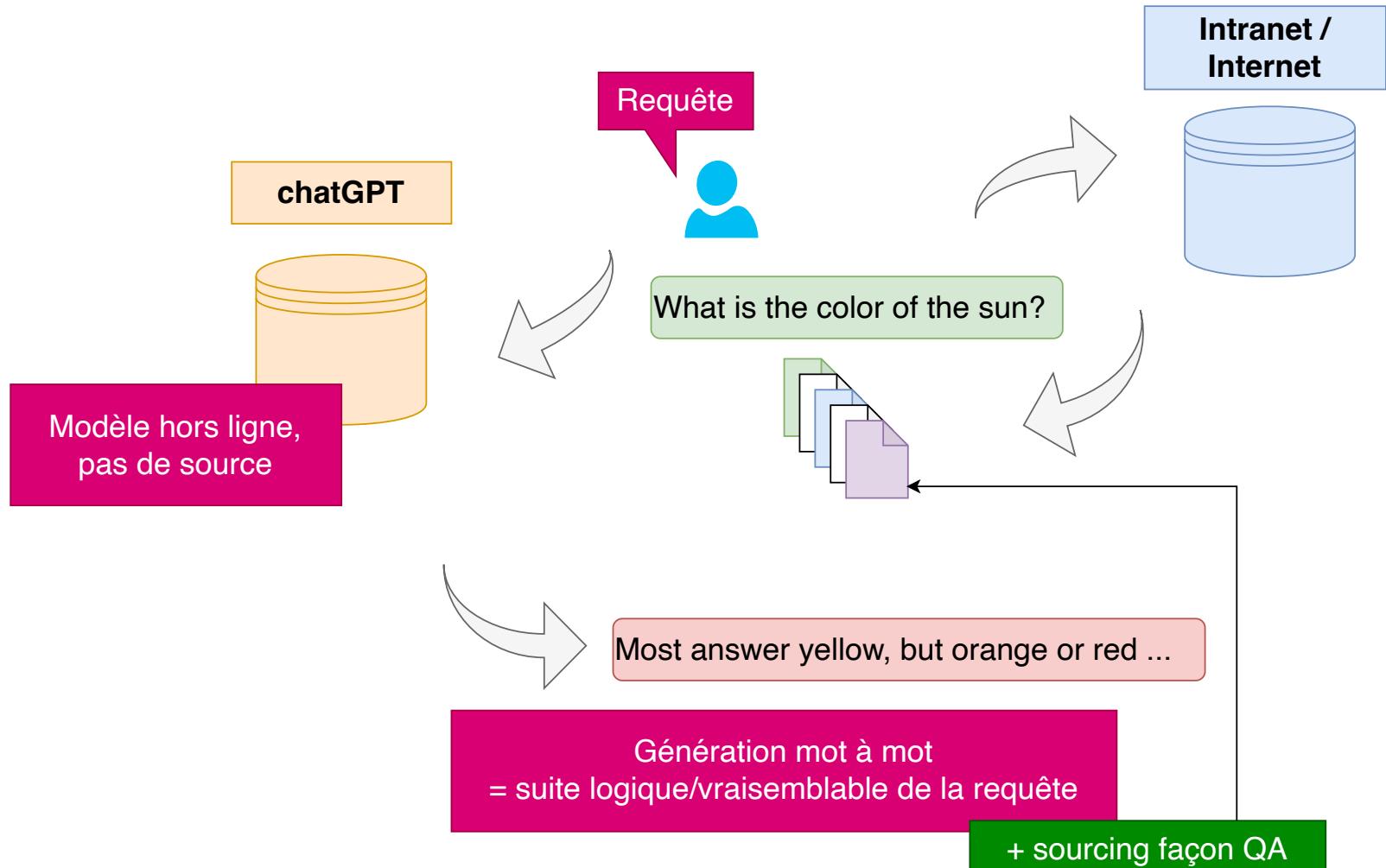
Paradigme: *Winner takes all*



- Architecture de type *Question Answering*
- Analyse de documents / résumés
- Comptes-rendus de réunion

Vers des systèmes combinés (RAG)

- Recherche d'information web + recherche de documents
 - Construction des réponses
- RAG : *Retrieval Augmented Generation*



→ Bing-ChatGPT vs. Bard



→ Des analyses comparatives des résultats

(ex : https://www.frandroid.com/culture-tech/intelligence-artificielle/1690803_google-bard-vs-bing-chat-vs-chatgpt-quel-est-le-meilleur-chatbot-par-ia)

- Pas d'explication sur l'intégration de ChatGPT pour Bing, de la construction des modèles, etc...
- Pas d'évaluation expérimentale
- Pas de rapport technique ou d'articles scientifiques

Plug-ins chatGPT

→ Applications tierces

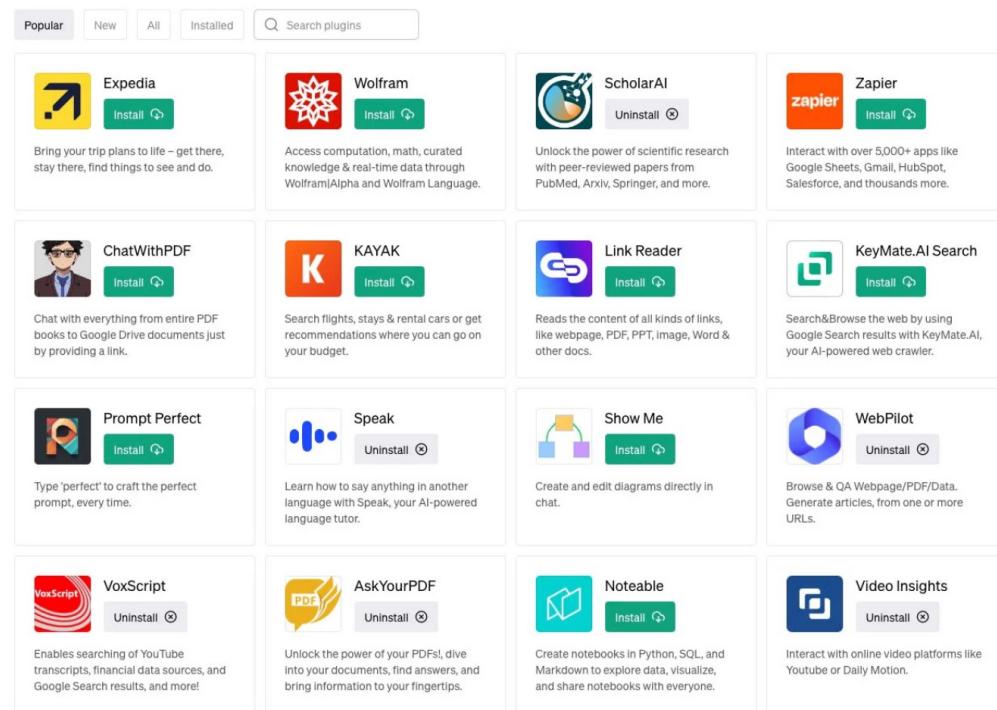
- ChatGPT intégré à part entière dans l'application
- Souvent une collaboration avec OpenAI

→ Plugins compatibles

- Permettre aux utilisateurs d'utiliser ChatGPT sur leurs applications

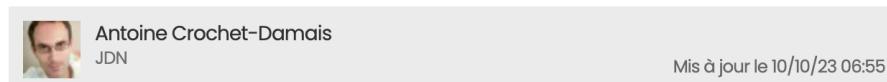
→ Plugins pour le SEO (Search Engine Optimization)

- Générer du contenu
- Analyser les pages web



- Focalisation sur les typologies d'usage
- Plus de détail dans la démonstration interactive à suivre

Plus 25% de rapidité, 12% de productivité et 40% d'amélioration : ChatGPT booste les performances des consultants... surtout les pires



L'enquête a été réalisée par Harvard et le MIT. Ils se sont penchés sur l'efficacité des consultants du Boston consulting Group, avec ou sans ChatGPT.

Selon une étude dirigée par Havard et le MIT, les consultants du Boston Consulting Group (BCG) voient la qualité de leur travail exploser de 40% avec l'utilisation de ChatGPT. L'étude en question a été réalisée auprès de 758 consultants de l'entreprise de conseil américaine. Un groupe utilisant ChatGPT a été comparé à un groupe témoin n'ayant pas recours à l'IA conversationnelle d'OpenAI. Le différentiel de 40% est mesuré au regard du niveau de qualité du travail fourni (voir le graphique ci-dessous).

Chapitre 8 : GPT4 et la multi-modalité

V



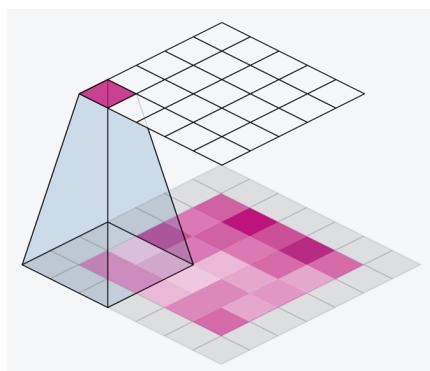
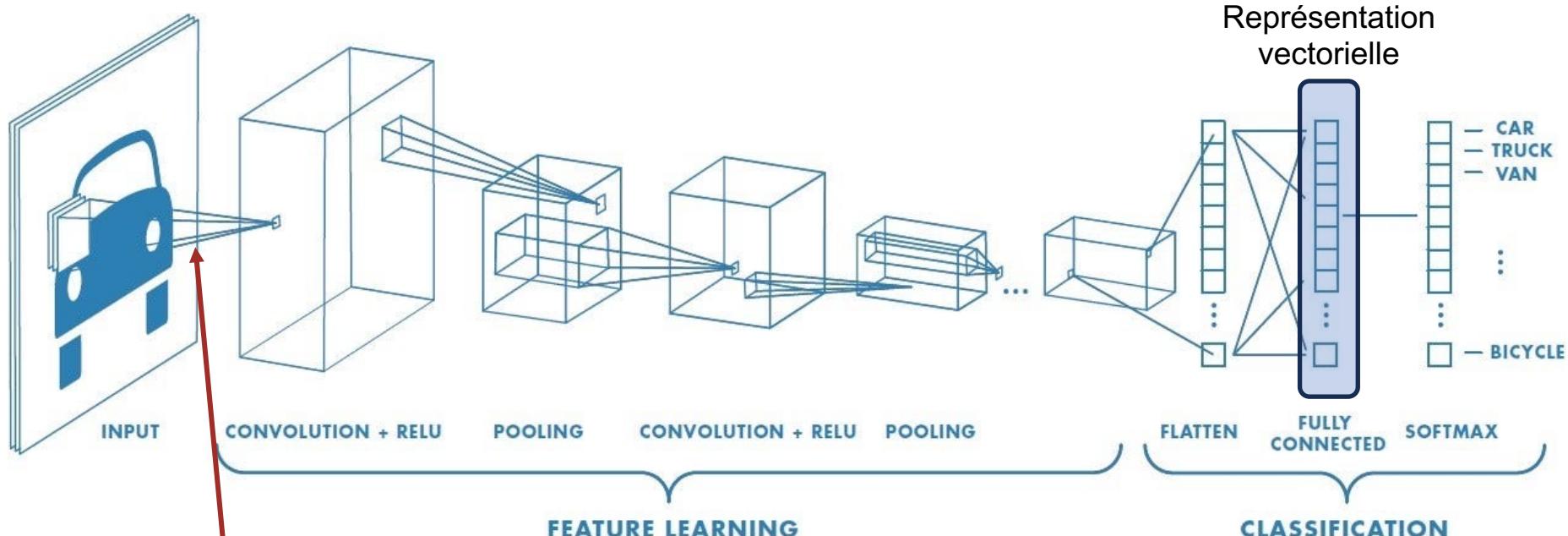
Peux tu me décrire ce bâtiment?



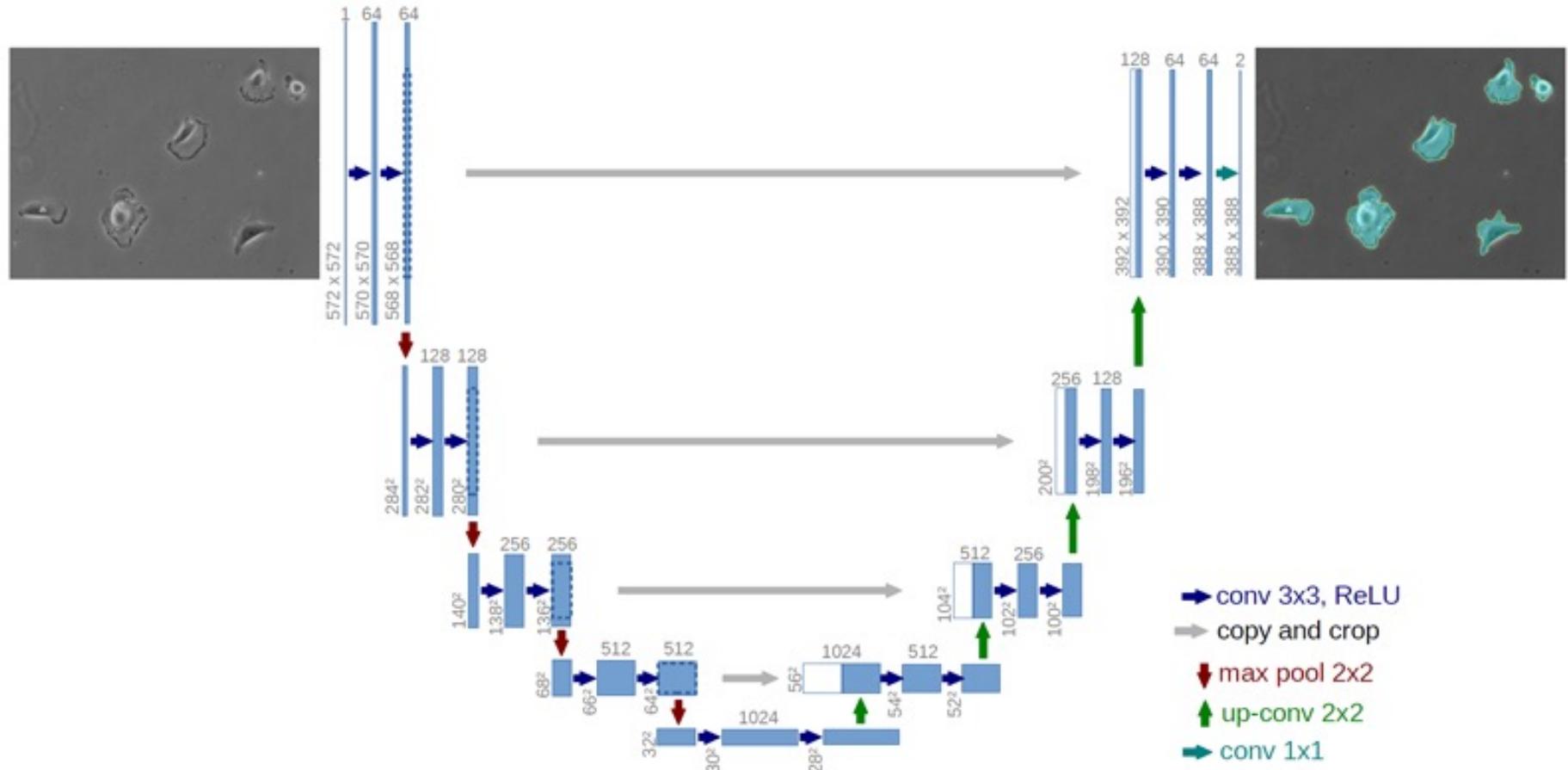
Ce bâtiment est la Casa Rosada, la résidence officielle du président de l'Argentine. Situé à Buenos Aires, il se distingue par sa couleur rose caractéristique. Le style architectural est éclectique, avec des éléments qui pourraient être classifiés comme relevant du néoclassique, avec ses colonnes et ses arcs, ainsi que du style italien, notamment avec ses ornements et détails.

Le bâtiment a plusieurs dômes et est richement décoré, avec des balustrades et des fenêtres ornées. On peut également apercevoir le drapeau argentin flottant au-dessus, symbolisant la fonction gouvernementale du lieu. C'est un bâtiment emblématique et historique, souvent associé aux grandes décisions politiques de l'Argentine et à divers événements historiques. La Casa Rosada est aussi un lieu touristique populaire à Buenos Aires, connu mondialement.

→ Extraire une représentation numérique d'image



- Systèmes symétriques :
- représentation générale => raffinement de l'image de sortie



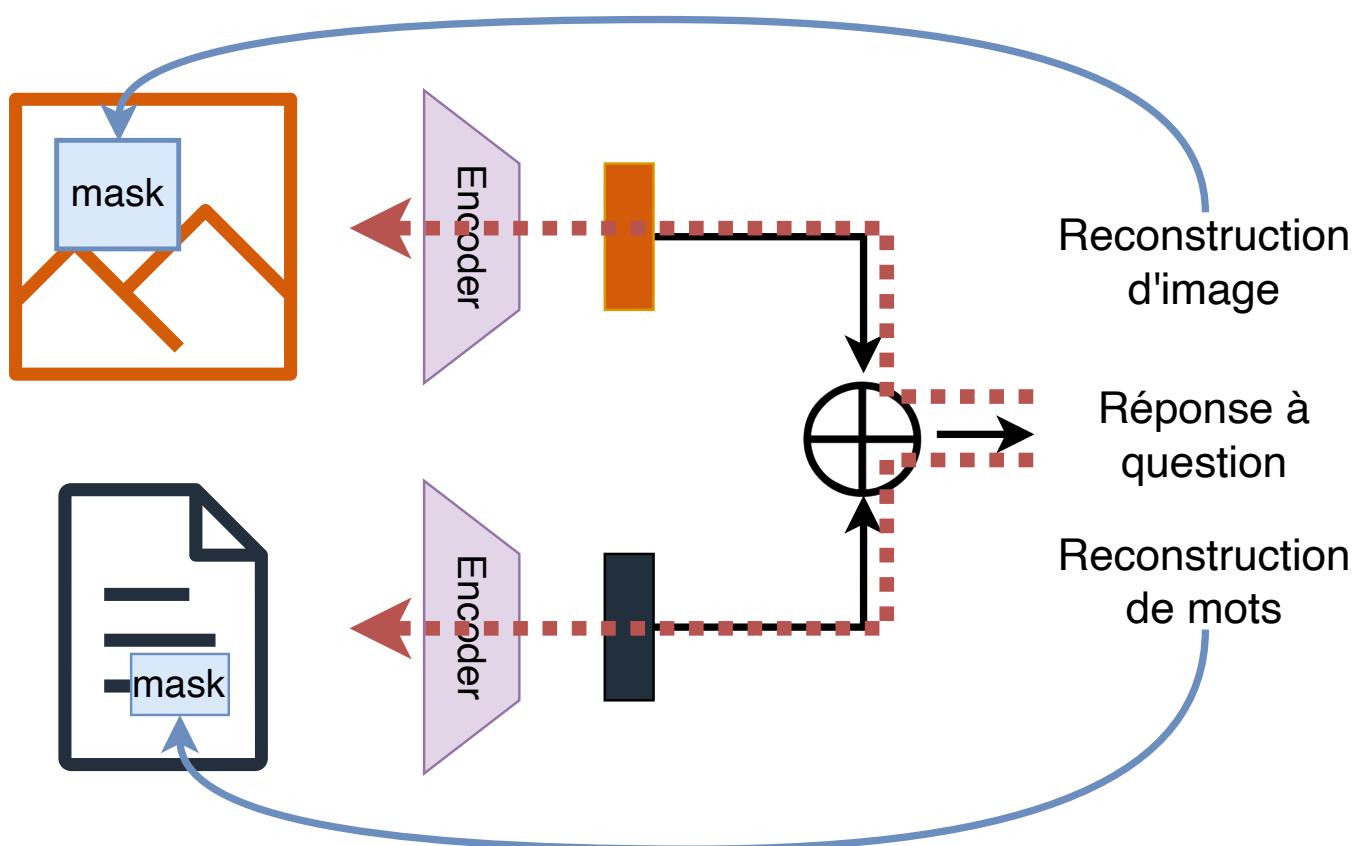
Biais des données textuelles

<i>Word</i>	<i>Teraword</i>	<i>Knext</i>
Spoke	11,577,917	372,042
Laughed	3,904,519	179,395
Murdered	2,843,529	16,890
Inhaled	984,613	5,617
Breathed	725,034	41,215

<i>Word</i>	<i>Teraword</i>	<i>Knext</i>
Hugged	610,040	11,453
Blinked	390,692	21,973
Was late	368,922	31,168
Exhaled	168,985	4,052
Was on time	23,997	14

- Beaucoup de non-dits
- *Foundation model* = encoder plusieurs modalités pour plus de généralité

1. Encodages des modalités => vecteurs
2. Mixages des représentations
→ addition, multiplication, concaténation, ...



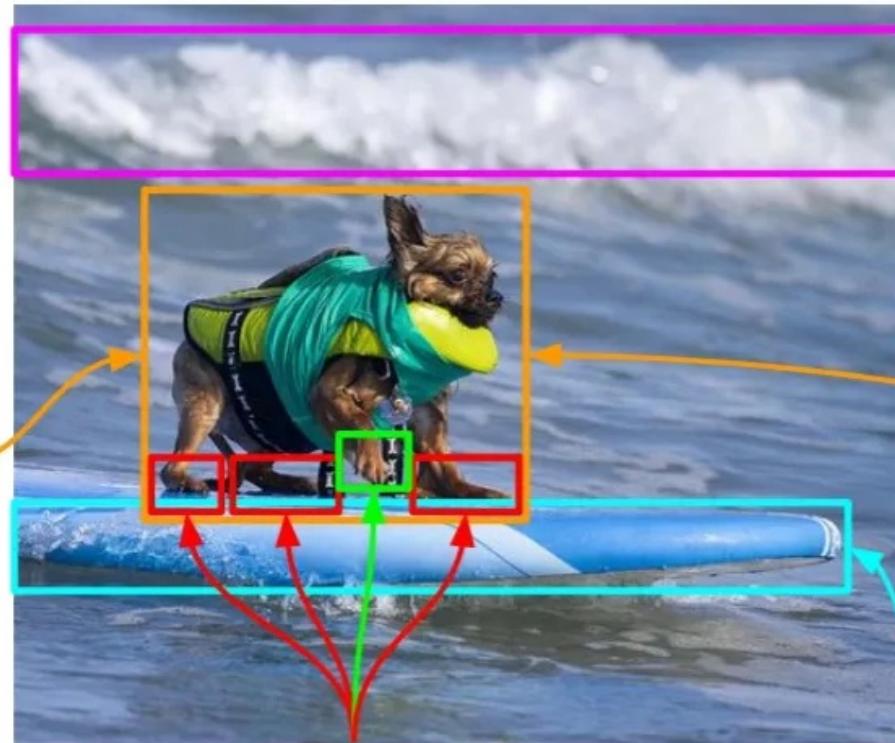
- **Grounding** = mêler texte et image
 - Nouvelles représentations des mots et des images

Where does this scene take place?

- A) In the sea. ✓
- B) In the desert.
- C) In the forest.
- D) On a lawn.

What is the dog doing?

- A) Surfing. ✓
- B) Sleeping.
- C) Running.
- D) Eating.



Why is there foam?

- A) Because of a wave. ✓
- B) Because of a boat.
- C) Because of a fire.
- D) Because of a leak.

What is the dog standing on?

- A) On a surfboard. ✓
- B) On a table.
- C) On a garage.
- D) On a ball.

Which paw is lifted?

Mieux comprendre le sens des mots

Différentes tâches multi-modales (2)

→ **Captioning** = générer une description textuelle de l'image



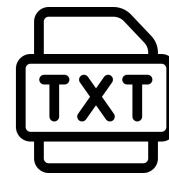
A cute little **dog** **sitting** in a heart drawn on a sandy **beach**.



A **dog** walking **next to** a little **dog** on top of a **beach**.



Numerical representation



Generative approach 96

Différentes tâches multi-modales (3)

→ Visual Question Answering (VQA) =
répondre à une question textuelle sur une image



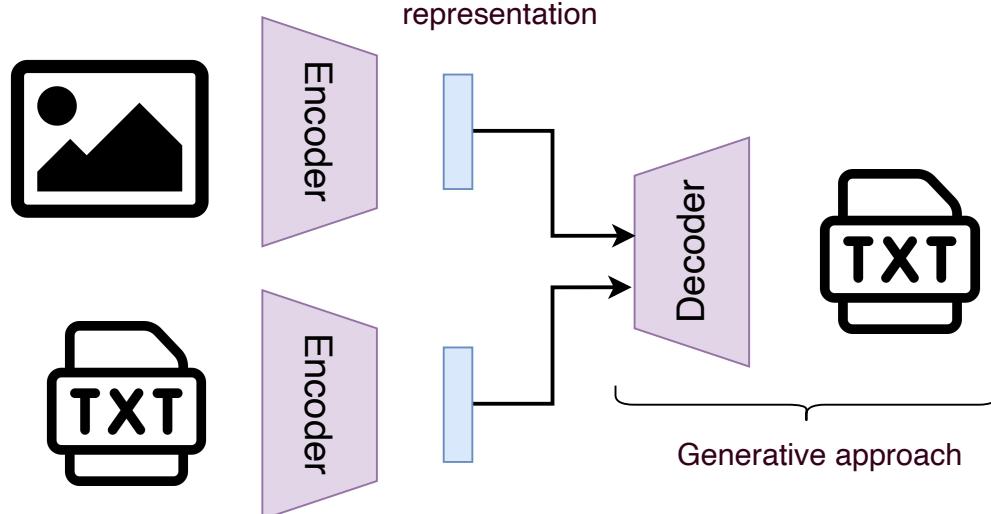
What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?



Différentes tâches multi-modales (4)

→ Gestion de factures

INVOICE

Your name/business name

Your address

Your phone number

Your email

Bill to: Invoice number: ######

Buyer name/business name Invoice date: MM/DD/YYYY

Buyer address Payment due: MM/DD/YYYY

Buyer phone number

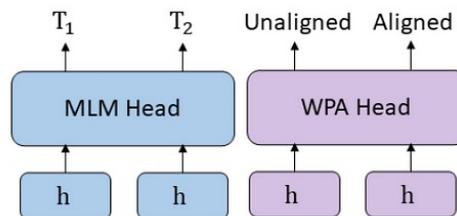
Buyer email

Item	Quantity	Price per unit	Amount
Item 1	#	\$0.00	\$0.00
Item 2	#	\$0.00	\$0.00
Item 3	#	\$0.00	\$0.00
		Subtotal	\$0.00
		Tax 0.00%	\$0.00
		Fees/discounts	\$0.00
		TOTAL	\$0.00

Terms and conditions

Terms and conditions go here

Pre-training Objectives

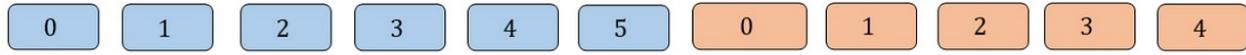


Multimodal Transformer

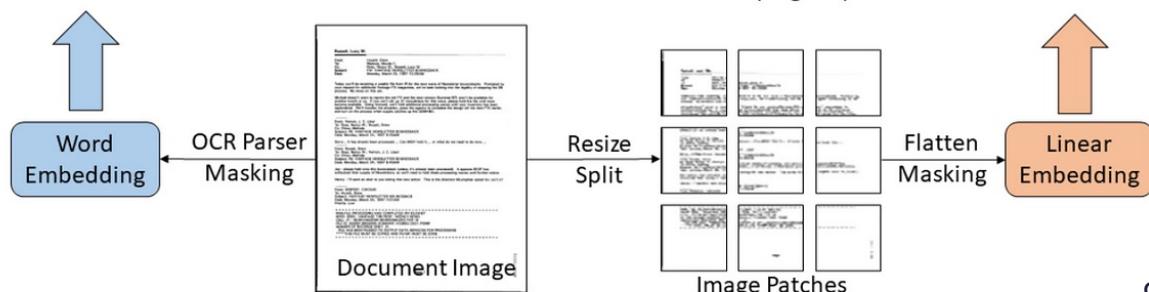
2D Position Embedding



1D Position Embedding



Word/Patch Embedding



Différentes tâches multi-modales (5)

→ Génération d'images: dall-E, midjourney,...

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research

as a 1990s Saturday morning cartoon as digital art in a steampunk style



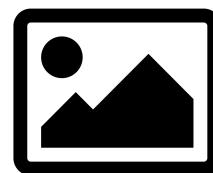
DALL-E 2



Numerical representation



Decoder



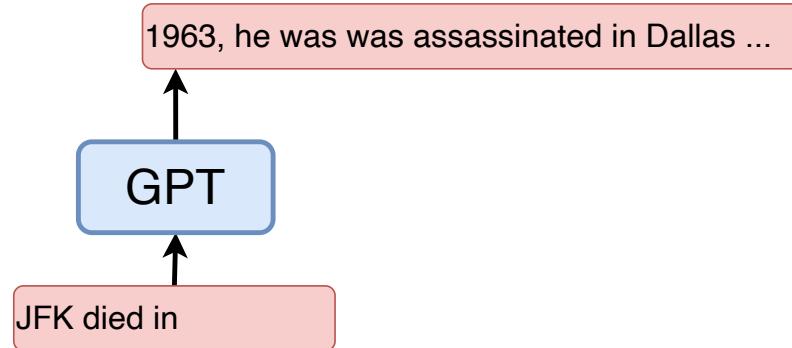
Generative approach

Chapitre 9 : Limites générales du machine learning

- Manque de **véracité / fiabilité**
 - Excellentes performances en moyenne... Mais des erreurs bêtes.
 - Quelle chance de faire une erreur sur la prochaine question?
 - Vérité vs Vraisemblance - Génération d'*hallucinations*
- Manque de **stabilité/prédicibilité**
 - Borner les comportements
 - *How old is Obama VS how old is Obama?*
- Manque d'**explicabilité/interprétabilité**
 - Pourquoi cette réponse?

→ Maximiser la **vraisemblance**

- respect de la grammaire, des accords,
- de la sémantique générale (un lieu pour un lieu, une date pour une date...)
- connaissances répétées = grammaire



→ Vraisemblance vs véracité

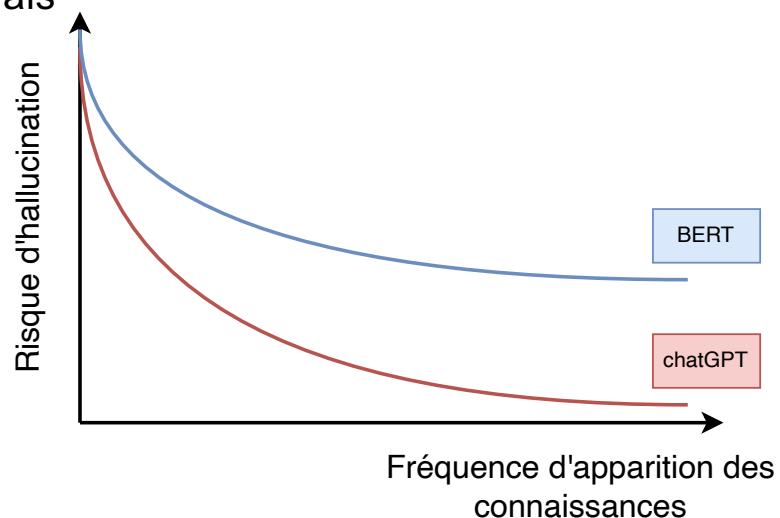
- Génération d'**hallucination** = mot crédible... Mais faux
 - e.g.: mauvaise personne, lieu inventé...

→ Fonctionnement hors ligne

- Pas de sources

→ Des réponses brillantes... Et des erreurs bêtes

- imprévisible



Modèle de langue = probabilité conditionnelle (très complexe)

- Jouer sur les connaissances rares
 - Bibliographie,
 - Entités dans les domaines spécifiques (bactéries, ...)

V Peux tu me donner une bibliographie des VAE sur les séries temporelles? 

Je m'excuse pour la confusion. Voici donc une sélection de références sur les Variational Autoencoders (VAE) pour les séries temporelles :

1. "Variational Autoencoder for Deep Learning of Images, Labels and Sequences" par Zhao et al. (2017) - Cette étude présente une approche VAE pour la modélisation de séquences temporelles en utilisant une architecture de réseaux de neurones récurrents.
2. "Variational Recurrent Autoencoders" par Chung et al. (2016) - Cette étude introduit une approche VAE pour la modélisation de séquences temporelles en utilisant des réseaux de neurones récurrents et une pour la phase d'entraînement.
3. "Generative Modeling for Time Series:

Variational recurrent auto-encoders

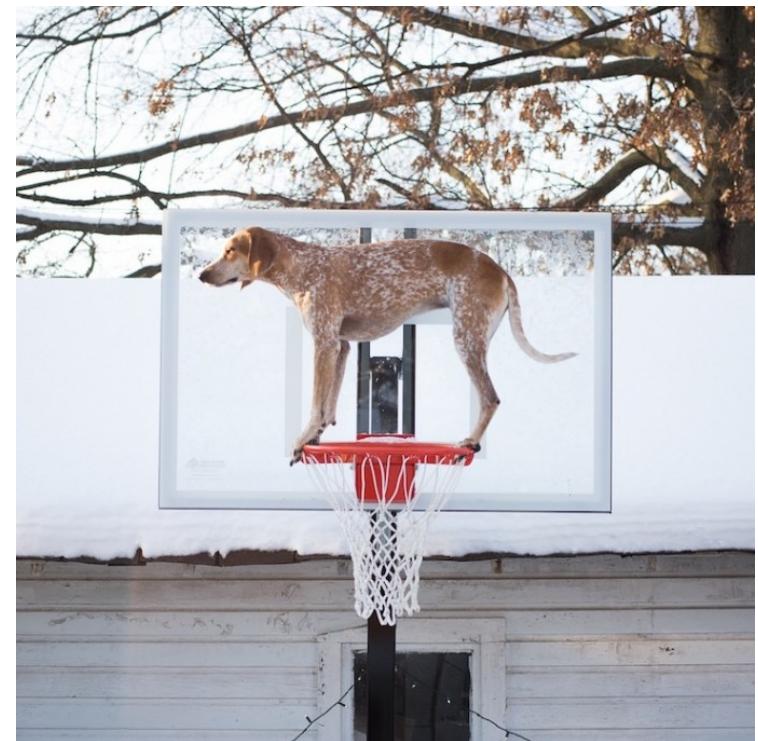
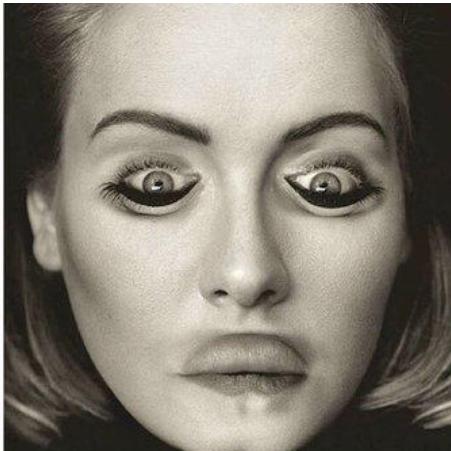
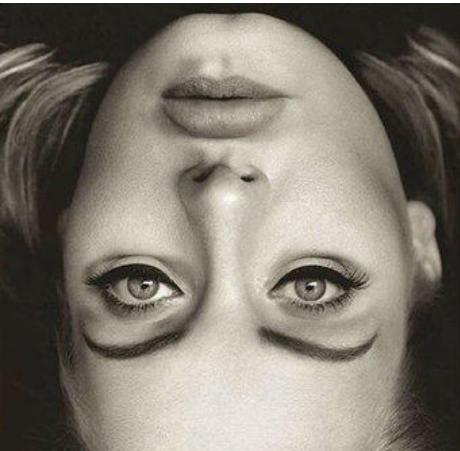
O Fabius, JR Van Amersfoort - arXiv preprint arXiv:1412.6581, 2014 - arxiv.org

In this paper we propose a model that combines the strengths of RNNs and SGVB: the **Variational Recurrent Auto-Encoder** (VRAE). Such a model can be used for efficient, large scale ...

☆ Enregistrer ⚡ Citer Cité 302 fois Autres articles Les 2 versions ➞

→ Questionner les diagnostics avec une représentation du monde

Faut-il simplement détecter un chien?
Est-il satisfaisant de reconnaître Adèle?



- Difficile de borner un comportement
- Impossible de prédire les bonnes/mauvaises réponses
 - Confiance du modèle décorrélée des erreurs !
- Petites modifications => grands changements

V how old is Obama

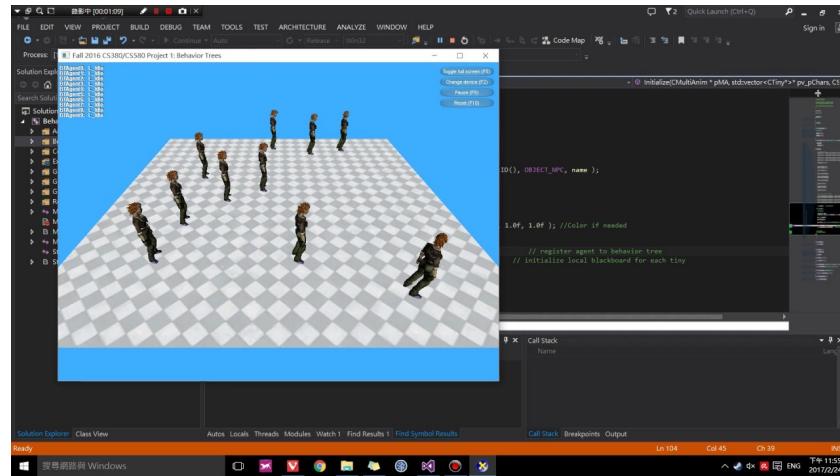
V how old is obama?

 As of 2021, Barack Obama was born on August 4, 1961, so he is 60 years old. thumb up thumb down

V and today?

 As a language model AI, I don't have real-time access to current dates. However, Barack Obama was born on August 4, 1961, so if today's date is after August 4th, 2021, he would be 60 years old or older. thumb up thumb down

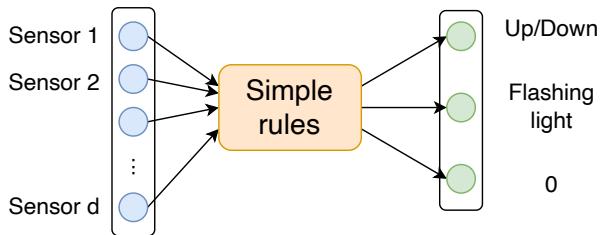
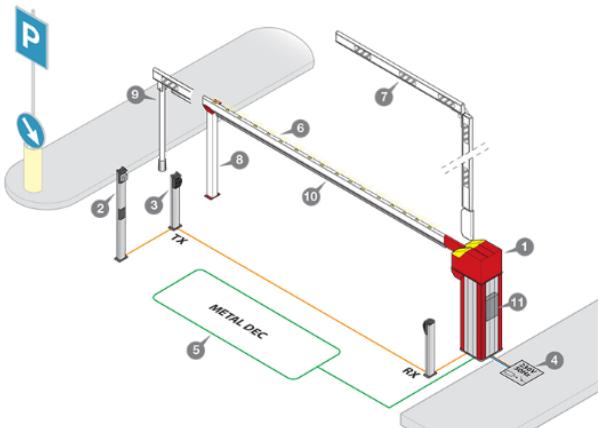
- Arbre/règles de comportement
 - monopole dans les jeux vidéos...
 - ... Et les outils de simulation



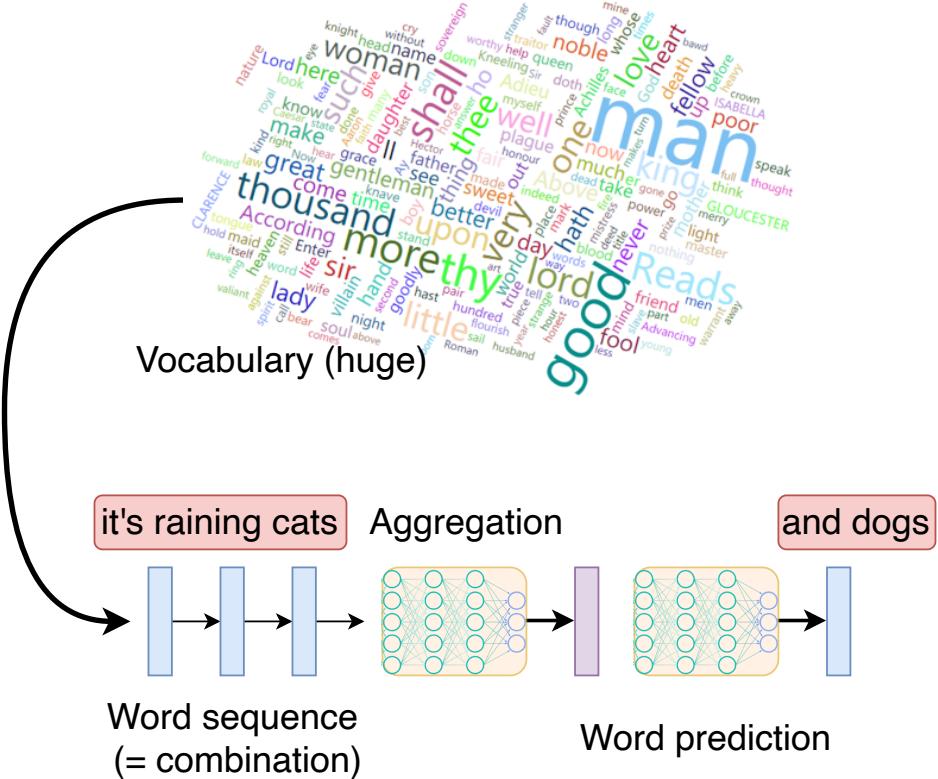
- Avantages:
 - Cout proportionnel à la complexité
 - Prouvable
 - Les comportements aberrants sont connus !
 - Déterminer ce que les agents voient ou ne voient pas



Explicabilité, stabilité et complexité



- Système *simple*
- Test exhaustif des entrées/sorties
- Comportement **prédictible & explicable**



- Grande dimension
- Combinations non-linéaires complexes
- Comportement **non-prédictible** & **non-explicable**

Explicabilité

→ Manque d'explicabilité

→ Modèles souvent perçus comme boîtes noires

→ Explicabilité

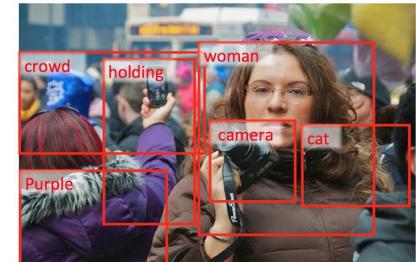
→ **a posteriori**

via des heatmaps,
analyse des causes de l'accident Uber

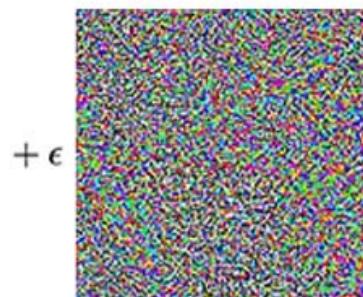
→ **intrinsèque**

via les systèmes à base de règles,
borner les comportements

→ **indirecte** via la génération explicite



→ La complexité est une source d'attaque



“panda”
57.7% confidence

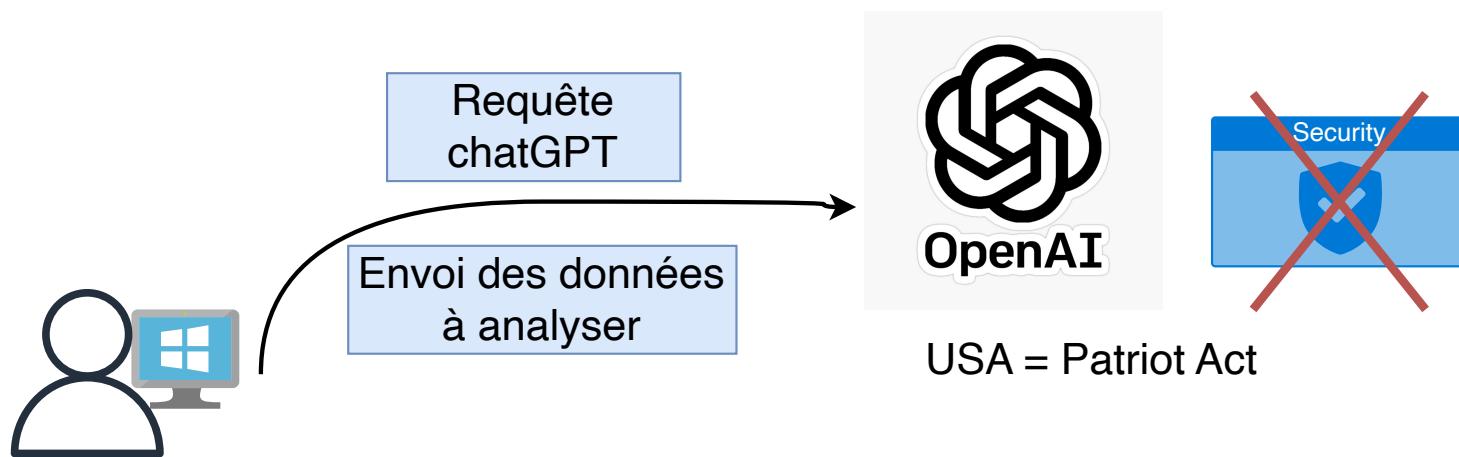
“gibbon”
99.3% confidence



- Taille des modèles de langue
 - Possibilité de cacher des logiciels à l'intérieur
- Exécution des modèles
 - Risque d'exécution de code / plug-ins
 - Risque de générer des malwares dans l'entreprise

- Manque de **véracité / fiabilité**
 - Excellentes performances en moyenne... Mais des erreurs bêtes.
 - Quelle chance de faire une erreur sur la prochaine question?
 - Vérité vs Vraisemblance - Génération d'*hallucinations*
- Manque de **stabilité/prédicibilité**
 - Borner les comportements
 - *How old is Obama VS how old is Obama?*
- Manque d'**explicabilité/interprétabilité**
 - Pourquoi cette réponse?
- Manque de **transparence**
 - Code privé, pas encore d'article scientifique
 - Modèle non externalisable
- Limite dans la **dimension des textes**
 - Limite à 2000 mots en entrée (32000 avec GPT-4)
- Un système que **ne sait pas tout faire**

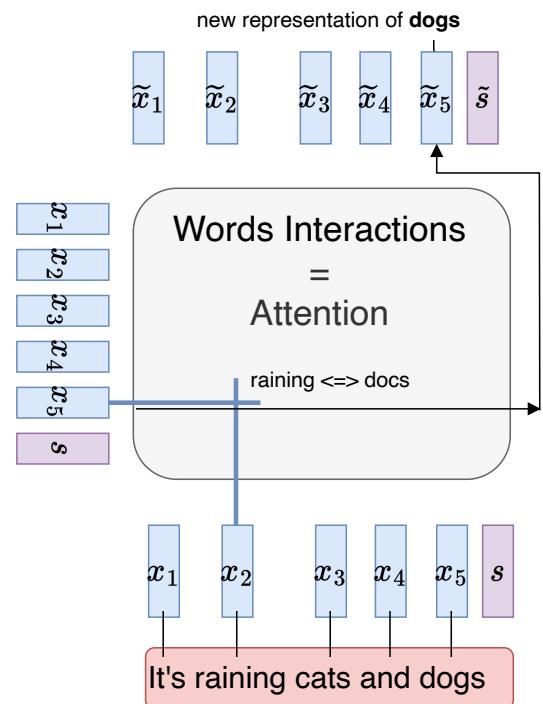
- ChatGPT = modèle non publié
 - Problème de transparence (pour les chercheurs)
- Données d'entraînement non publiées
 - Problème d'évaluation des performances
- Modèle uniquement disponible chez OpenAI/Microsoft
 - non externalisable
 - Impossible de traiter des données sensibles (médicales, entreprises, ...)



Coût machine et limite de la taille des entrées

- Analyse des interactions de mots complexité quadratique
- Multiplication des paramètres

- ⇒ Limite dans la taille des textes analysés
- Modèle de langue 2022 typique: 512 tokens
 - ChatGPT: 2000 tokens
 - GPT4: 32k tokens



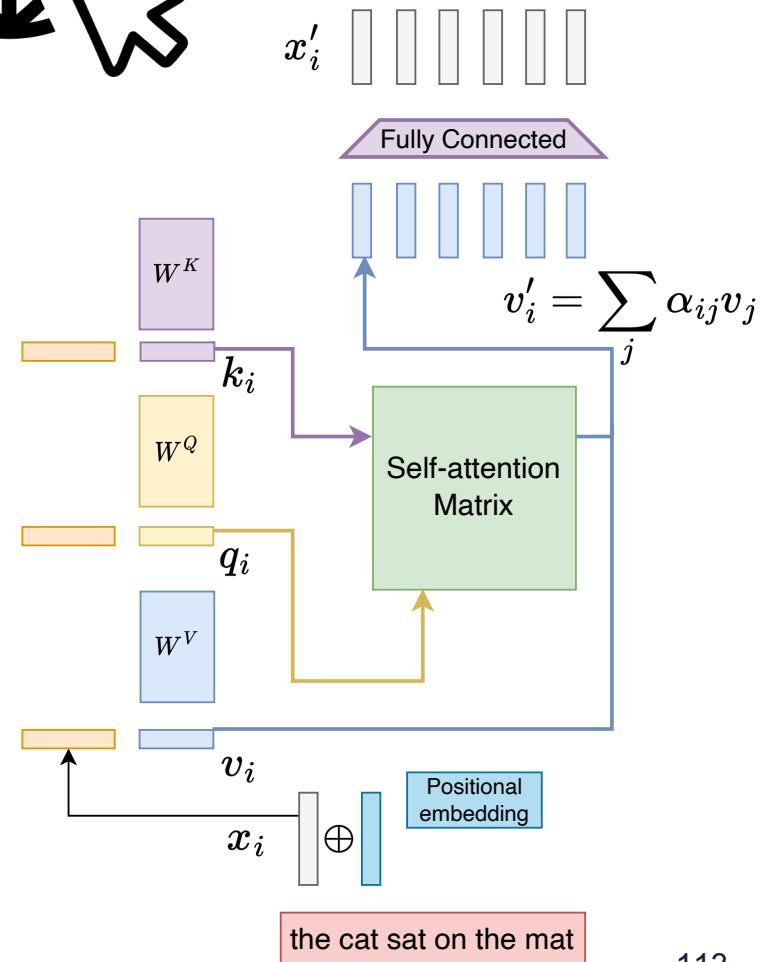
L'augmentation de la capacité à un coût très important

~500 tokens / page => 60 pages au maximum



- En apprentissage:
 - Voir des milliards de phrases
 - Mettre à jour des milliards de paramètres

- En inférence:
 - Beaucoup de calcul pour découvrir le mot suivant (et le suivant...)



the cat sat on the mat

Ce qu'il ne sait pas faire

→ Parce qu'il n'a pas été entraîné

V

Peux tu me donner 10 animaux en 4 syllabes?



Bien sûr ! Voici 10 animaux en 4 syllabes :

1. Éléphant
2. Crocodile
3. Alligator
4. Kangourou
5. Dinosaur
6. Chimpanzé

V

peux tu calculer 4012000 divisé par 78000 ?



Bien sûr ! 4012000 divisé par 78000 égale $51,5$.

V

et 4012 divisé par 78 ?



4012 divisé par 78 égale environ $51,48717949$ (arrondi)

>>> $4012 / 78$
51.435897435891

Chapitre 10 : Risques associés aux modèles de langue

Typologie des risques



Discrimination, exclusion and toxicity

Harms that arise from the language model producing discriminatory and exclusionary speech.



Information hazards

Harms that arise from the language model leaking or inferring true sensitive information.



Misinformation harms

Harms that arise from the language model producing false or misleading information.



Malicious uses

Harms that arise from actors using the language model to intentionally cause harm.



Human-computer interaction harms

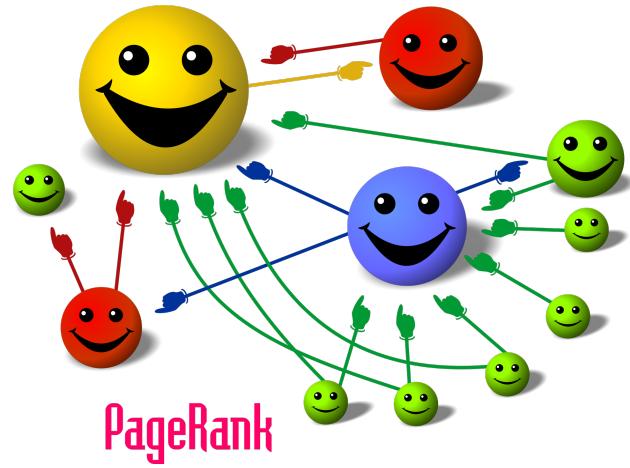
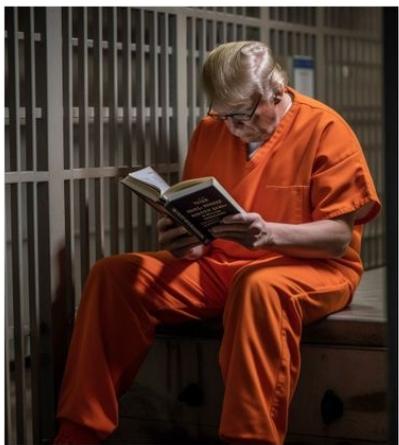
Harms that arise from users overly trusting the language model, or treating it as human-like.



Automation, access and environmental harms

Harms that arise from environmental or downstream economic impacts of the language model.

- IA générative = possibilité de générer des masses de fausses informations
 - Images, textes
- L'autorité des informations est actuellement liée :
 - Nature : inconsciemment, image = vérité
 - Source : journaux, réseaux sociaux, ...
 - Masse : nombre de variantes, de citations (pagerank) etc...



- Harcèlement en ligne
- Vérification / Générer des rapports entiers
 - Comment détecter/corriger/analyser ces masses d'information

- Informations personnelles
 - Nouveaux outils d'accès à l'information
 - Remise en cause technique du droit à l'oubli
 - Nouvelle mémoire du web
- Informations interdites
 - Fabrication d'une bombe
- Informations immorales/illégales
 - apologie du nazisme, du racisme
- Informations blessantes
 - Sur le fond ou la forme



DROIT À L'OUBLI



- Risque d'anthropomorphisation de l'algorithme
 - Distinguer l'homme de la machine
 - Dans la vie quotidienne
 - A l'hôpital, avec les robots thérapeutes
- Dead-bot
 - Apprendre de données, apprendre un style
 - Répondre à la place des morts
- Comprendre/Expliquer les suggestions d'un système de recommandation



Besoin de transparence lorsqu'un algorithme répond

- Ce que les humains font :
 - Comportement collectif, émanant d'activité sociale et d'un monde commun (e.g., acquérir du langage)
- Ce que les modèles de langue font/sont :
 - Modèles mathématiques génératifs basés sur la distribution statistiques des mots
 - Peuvent être utilisés comme modèles génératifs et/ou encapsulés dans des systèmes plus grands

Différents besoins, différents contextes pour une même comportement :

la prédition du prochain mot

« The first person to walk on the Moon was » → Neil Armstrong

« Twinkle twinkle » → little star

« After the ring was destroyed, Frodo Baggins returned to » --> the Shire

→ Machine Learning = extraction de biais dans les données



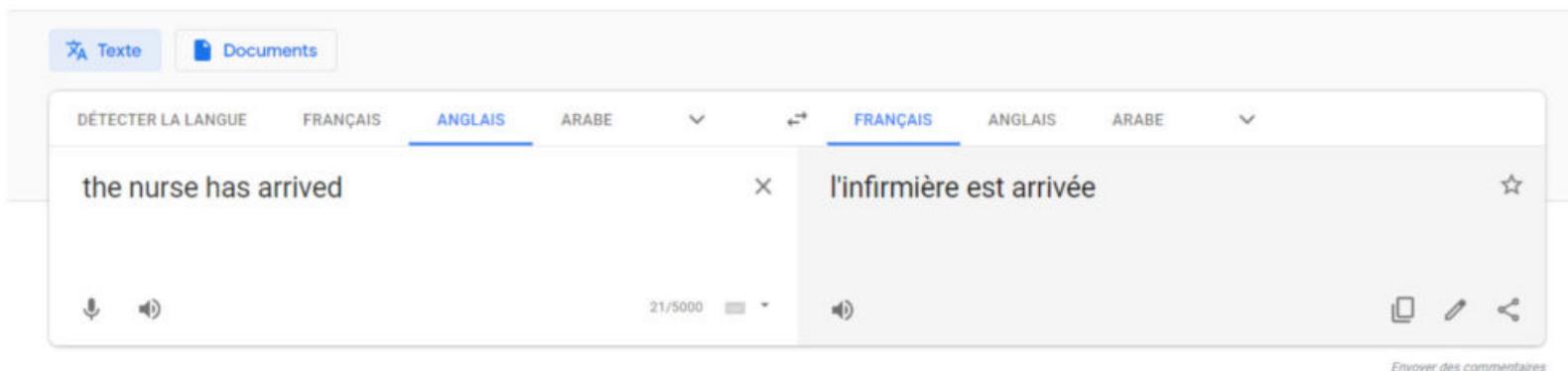
Oreilles pointues,
moustaches, texture de poils
=
Chat



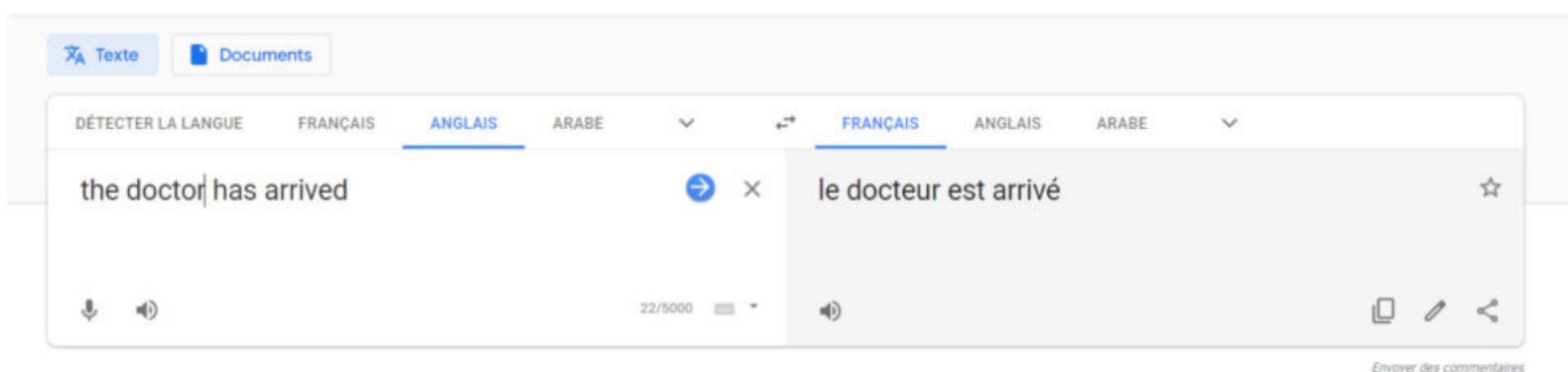
Homme blanc, +40ans,
costume
=
Cadre supérieur

Apprentissage sur des données **sélectionnées et passées** :
accentuation + reproduction de biais historiques

→ Quelle vision du monde à travers les textes que nous lisons ?



The screenshot shows a Google Translate interface. At the top, there are tabs for "Texte" and "Documents". Below that, language selection bars show "DÉTECTOR LA LANGUE" followed by "FRANÇAIS", "ANGLAIS", and "ARABE". The "ANGLAIS" tab is selected. On the right, another set of language selection bars shows "FRANÇAIS" (selected), "ANGLAIS", and "ARABE". The input text "the nurse has arrived" is on the left, and the output translation "l'infirmière est arrivée" is on the right. Below the text boxes are microphone and speaker icons, a character count indicator (21/5000), and a progress bar. To the right of the output text are edit and share icons. A small link at the bottom right says "Envoyer des commentaires".



This screenshot shows another Google Translate interface. The layout is identical to the first one. The input text "the doctor has arrived" is on the left, and the output translation "le docteur est arrivé" is on the right. A blue circular arrow icon is next to the input text, indicating it's being processed. The bottom right corner has a link that says "Envoyer des commentaires".

- Biais dans les images:
 - Genre, couleur de peau, posture,...
- Directement lié aux données d'entraînement



Seule option: une correction manuelle

Biais dans les données



Colin Madland
 @colinmadland

...

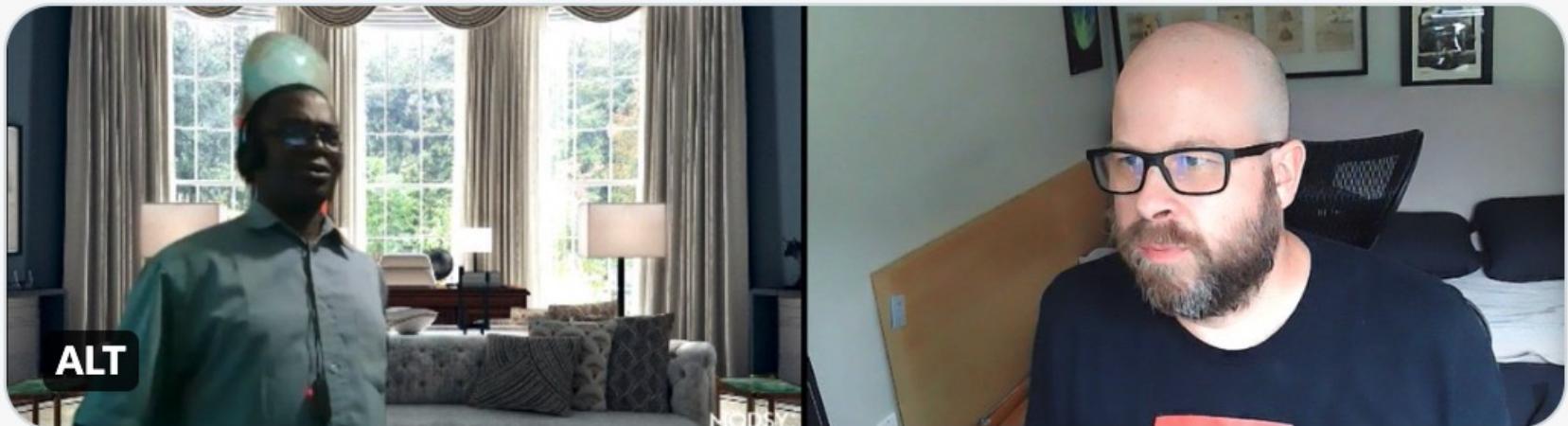
any guesses?





Colin Madland 🇺🇦 @colinmadland · 19 sept. 2020

...



- Correction des biais:
 - Sélection de données spécifiques, ré-équilibrage
 - Censure de certaines informations
 - e.g. origine ethnique en RH
 - Censure des résultats de l'algorithme
 - e.g. apologie du nazisme dans chatGPT



Correction des biais = travail éditorial

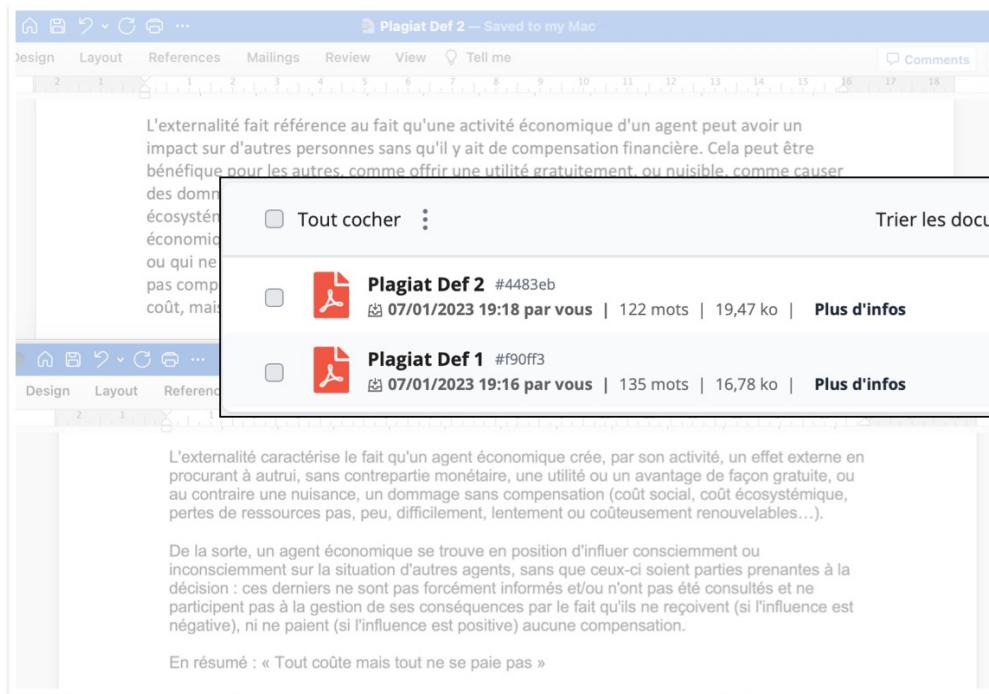
- Qui effectue ce travail?
 - Experts métiers lors de la rédaction d'un cahier des charges
 - Ingénieurs, lors de la conception des algorithmes
 - Groupe éthique, lors de la validation des résultats

Quelle légitimité?

- Comment faire?
 - Masquer des données est difficile =>
algorithme = reconstruction à partir des autres données

Quelle efficacité?

- Déetecter un texte wikipedia = signature d'un objet existant
- Quid des modèles de langue?



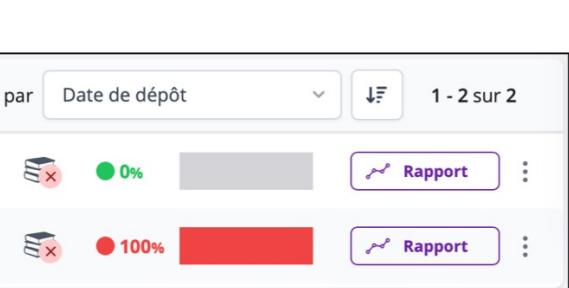
L'externalité fait référence au fait qu'une activité économique d'un agent peut avoir un impact sur d'autres personnes sans qu'il y ait de compensation financière. Cela peut être bénéfique pour les autres, comme offrir une utilité gratuitement, ou nuisible, comme causer des dommages à l'environnement.

L'externalité caractérise le fait qu'un agent économique crée, par son activité, un effet externe en procurant à autrui, sans contrepartie monétaire, une utilité ou un avantage de façon gratuite, ou au contraire une nuisance, un dommage sans compensation (coût social, coût écosystémique, pertes de ressources pas, peu, difficilement, lentement ou coûteusement renouvelables...).

De la sorte, un agent économique se trouve en position d'influer consciemment ou inconsciemment sur la situation d'autres agents, sans que ceux-ci soient parties prenantes à la décision : ces derniers ne sont pas forcément informés et/ou n'ont pas été consultés et ne participent pas à la gestion de ses conséquences par le fait qu'ils ne reçoivent (si l'influence est négative), ni ne paient (si l'influence est positive) aucune compensation.

En résumé : « Tout coûte mais tout ne se paie pas »

Reformulation par chatGPT



Document	Similarity Score	Plagiarism Status
Plagiat Def 2 #4483eb	0%	Plagiat
Plagiat Def 1 #f90ff3	100%	Plagiat

Définition de Wikipedia

Crédit: S. Pajak

Texte Wikipedia vs Texte reformulé par chatGPT

GPTZero

Detect AI Plagiarism. Accurately



Chat GPT



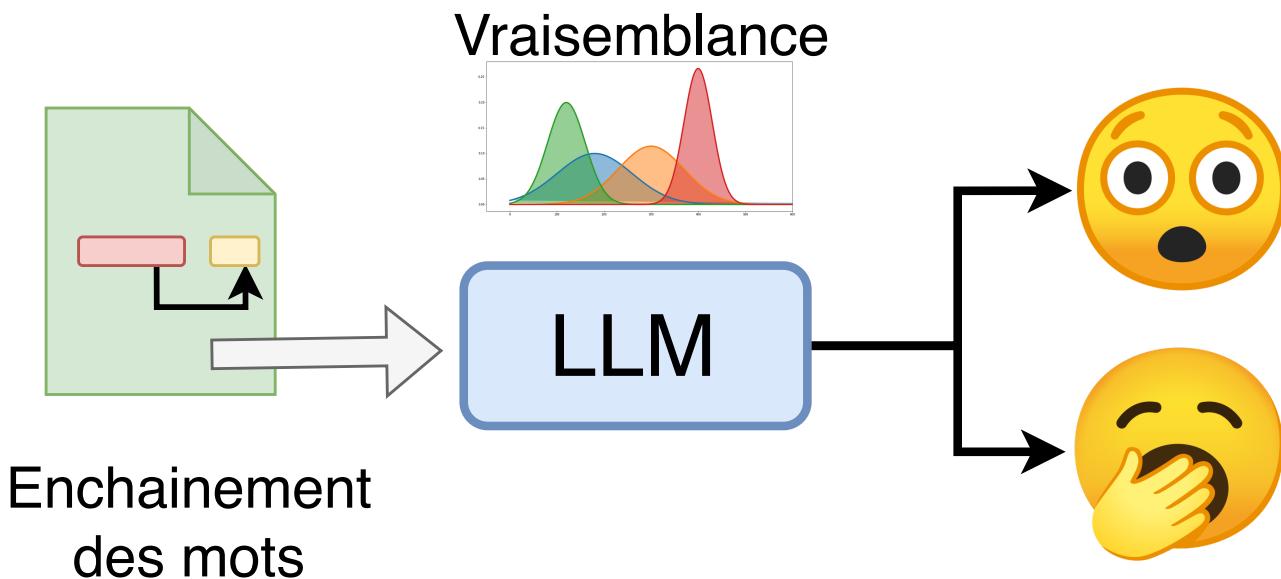
AI Detector

Torchbankz

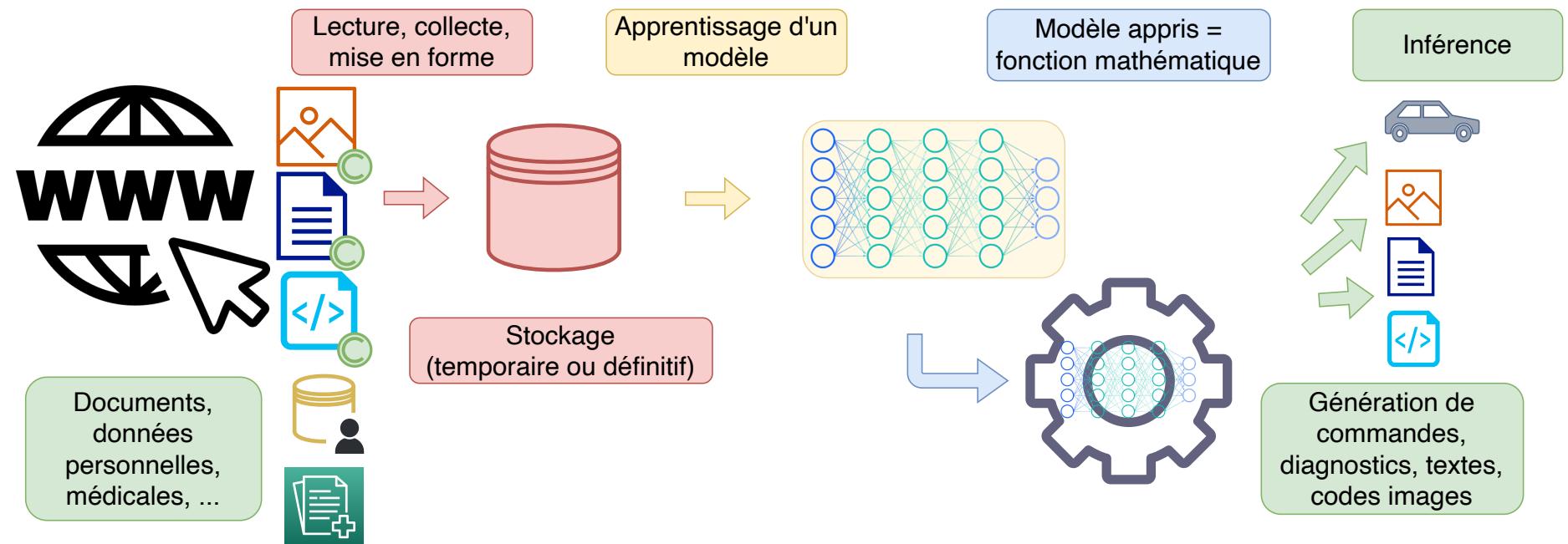
- DéTECTer les algorithmes génératifs
- Basé sur la vraisemblance des textes
- Actuellement efficace... Statistiquement!
 - Sur des **textes longs** : style, mots de liaison surabondants...
 - Pas efficace à 100%
 - Performance en baisse si reformulation automatique
 - Sur les artefacts des images

Aujourd'hui possible... Mais demain?

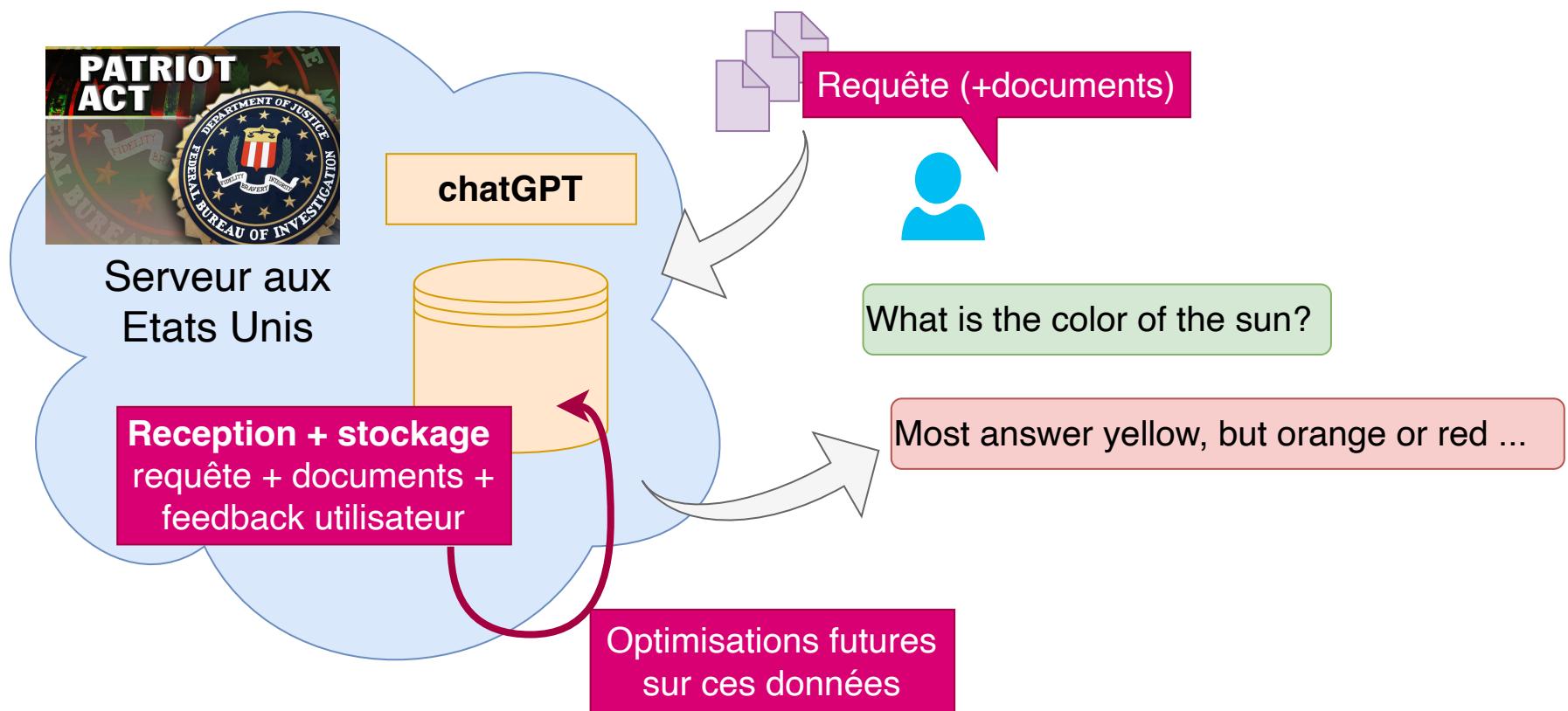
- Vraisemblance/surprise des enchainements de mots observés
- Quelles sources sont utilisées dans l'apprentissage d'un LLM?
 - Expliquer les performances
 - Repenser les questions de copyright



Questions juridiques

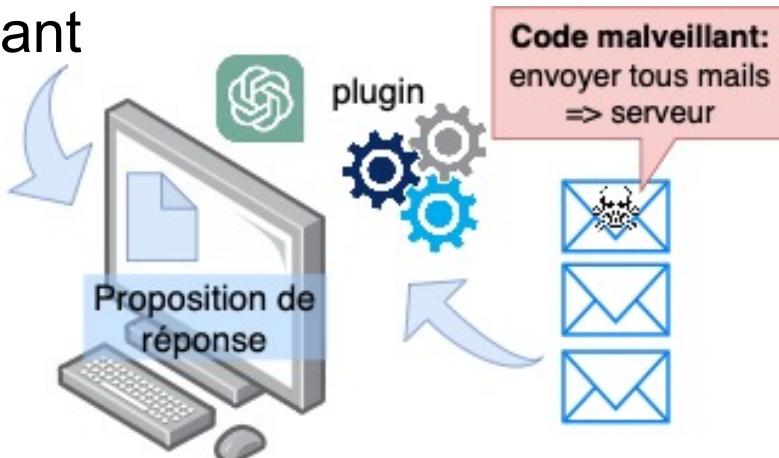


→ Usage de modèle externe = transfert de données



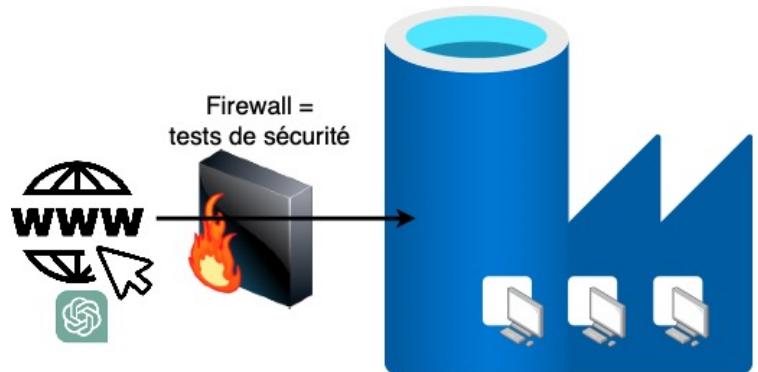
- Risque d'exécution de code malveillant

- Injection de code



- Récupération des LLM

- de plus en plus gros



- Risque de génération de code malveillant

- En masse => remise en cause

des systèmes de détection de signature

- A l'intérieur même de l'entreprise!



Chapitre 11 : Impacts sociaux

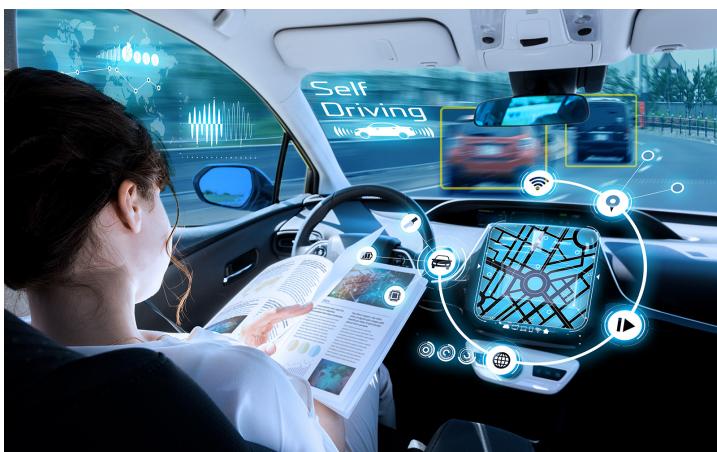
Impacts dans de nombreux domaines

Education



Source : <https://fr.unesco.org/themes/tic-education/intelligence-artificielle>

Mobilité



Source : <https://knaufautomotive.com/fr/la-voiture-autonome-quand-les-vehicules-naureront-plus-besoin-de-conducteur/>

Santé



Source : <https://www.echo sciences-grenoble.fr/communautes/le-magazine-des-sciences-de-rcf-isere/articles/l-intelligence-artificielle-et-la-medecine>

Justice



Source : <https://www.afis.org/Les-systemes-d-intelligence-artificielle-quelle-responsabilite>

(1) Ne pas se tromper de sujet / domaine

Numérisation / digitalisation \neq Intelligence Artificielle

- ▶ **Supermarché:** Automatisation des caisses, de l'étiquetage des prix...
- ▶ **Secrétariats délocalisés:** Doctolib, ...
- ▶ **Dématérialisation des courriers**
- ▶ **Service public:** Sites webs, FAQ, échanges numériques
- ▶ **Jeux vidéos**
- ▶ **Manutention:** Robotisation 80' \Rightarrow 2000'
- ▶ **Tous domaine:** Chatbot <2020'
- ▶ **Logistique:** Optimisation
- ▶ **Services en ligne:** Traduction, filtre spam, systèmes de recommandation
- ▶ **Publicité:** ciblage, personnalisation
- ▶ **Sécurité:** reconnaissance faciale
- ▶ **Manutention:** Robotisation >2010'
- ▶ **Tous domaine:** Chatbot >2020'

(2) Ne pas se tromper d'époque

IA aujourd'hui \neq IA demain

- ▶ Publicité en ligne

La publicité représente entre 80 et 90% du CA de Google

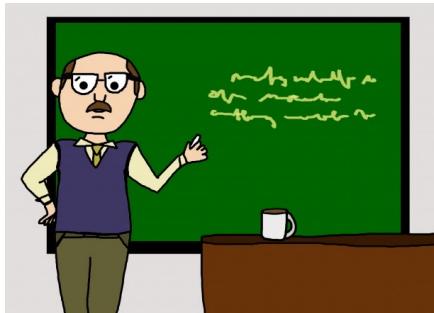
- ▶ Reconnaissance faciale
- ▶ Service en ligne : traduction, filtre spam, systèmes de recommandation
- ▶ Logistique
- ▶ Robotique (émergent)
- ▶ chatbot (émergent)

- ▶ Véhicule autonome
- ▶ Aide au diagnostic médical
- ▶ Traitement automatique des mails, CR réunion, rédaction des courriers
- ▶ Bio-informatique: molécule
- ▶ Nouvelle agriculture
- ▶ Conception de structures (optimisation matériaux/résistance...)
- ▶ Jeux vidéos? Robotique avancée?
- ▶ **Ressources Humaines**



- ChatGPT affirme que la terre est ronde, et non plate
- D'où vient cette information?
 - Dominante web?
 - Classement des sources par les ingénieurs dans les corpus?
 - Ligne éditoriale?
- Quelle régulation sur les lignes éditoriales des algorithmes?

La transparence est nécessaire... Mais difficile à mettre en place
Lier à la citation des sources



Pour les enseignants

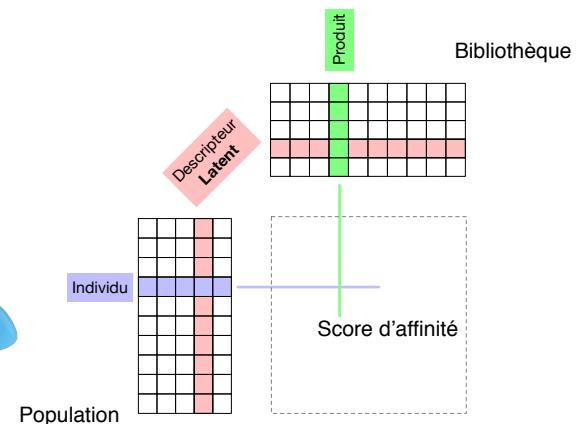
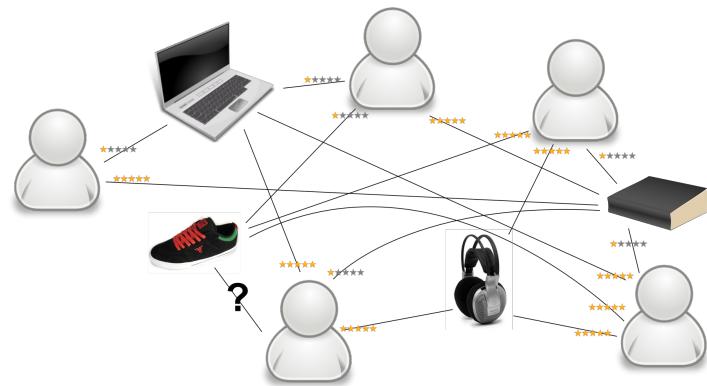


Pour les élèves/étudiants

- Création de ressources : génération libre ou contrôlée (par les enseignants) ?
- Nouveaux modes d'acquisition des connaissances : ressources variées, multimodales et « gamifiées » ?
- Création de parcours d'apprentissage : construction d'un profil utilisateur
- Amélioration de l'accès et de la découvrabilité des ressources : varier les ressources
- Evaluer, se préparer aux examens : créer et corriger des examens

Mais avant tout, un nouvel outil d'accès à l'information

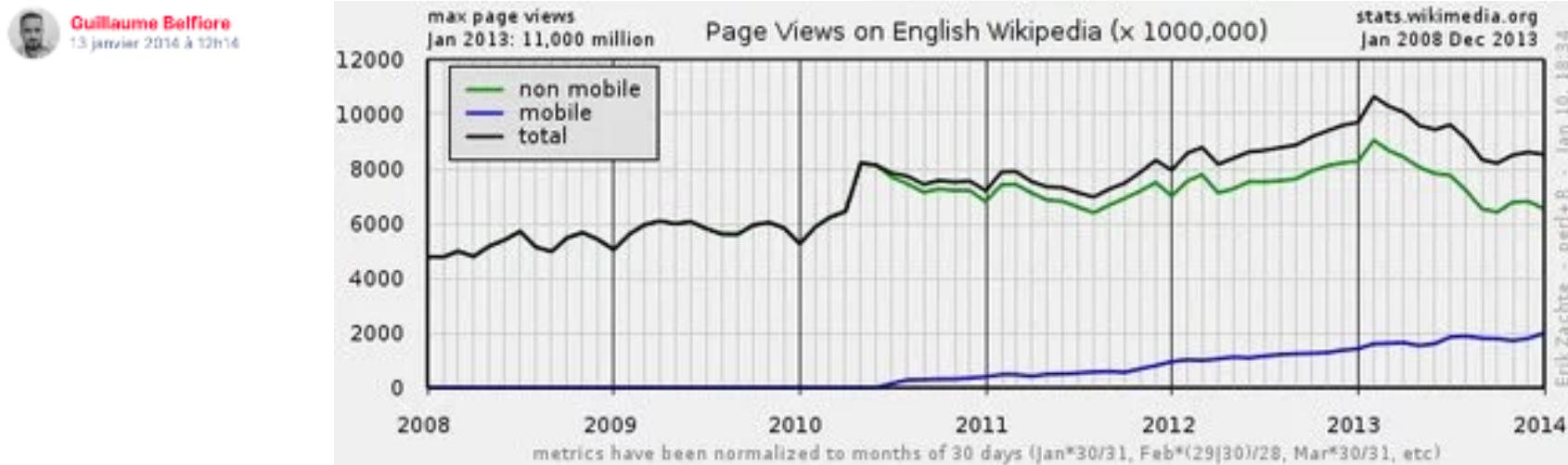
- Comprendre le comportement des apprenants et des professeurs
- Identifier des profils
 - Que construisent-ils ? Que consultent-ils ?
 - Comment apprennent-ils ?
- Potentiel :
 - Optimiser la génération/suggestion de contenu
 - Optimiser les espaces de vente
 - Fidéliser les apprenants et professeurs



Après l'ère des *interactions utilisateurs*, retour à l'analyse des données textuelles

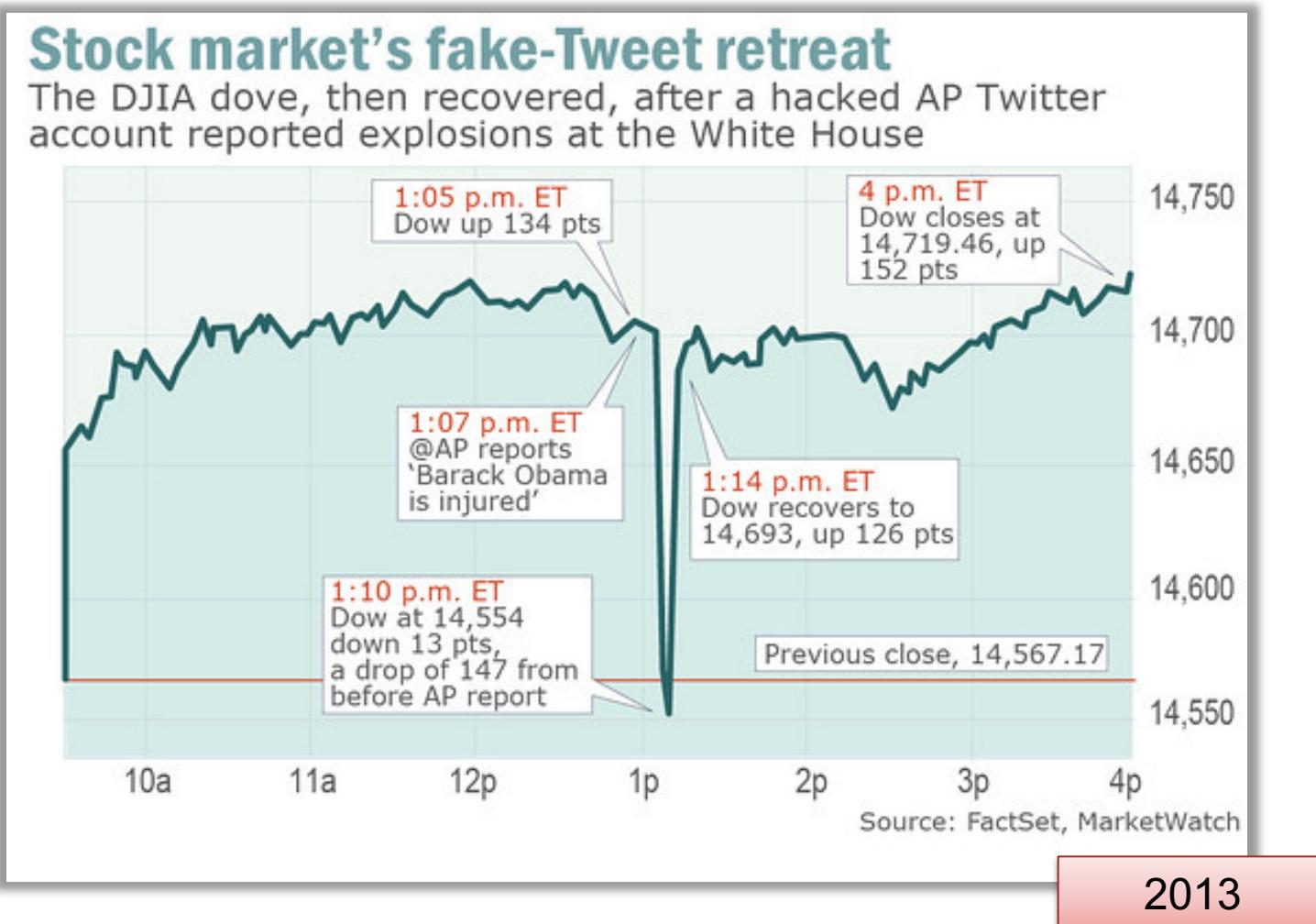
- Créeur de contenu = rémunération publicitaire = visite
- En 2013, la mise en place du Google Knowledge Graph a fait baisser significativement l'audience de wikipedia

Google Knowledge Graph aurait causé une baisse du trafic de Wikipedia en 2013



Quel modèle économique pour les sites de demain?

→ Analyse des données & trading



Séries temporelles & analyse des données textuelles

Chapitre 12 :

Conclusion

Vue géographique des gros modèles de langue

→ A l'international

- Des gros acteurs : OpenAI, GAMMA, ...
- BloombergGPT : version « finance » de ChatGPT



→ En France :

- Nombreux chercheurs et laboratoires de recherche
- Modèles de langue français : Camembert [1] (FAIR, INRIA et Sorbonne Université) et Flaubert [2] (Universités Grenoble Alpes, de Paris Diderot et PSL)



CamemBERT
A Tasty French Language Model

- Supercalculateur Jean Zay (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation - GENCI)

- 36 millions de milliards d'opérations par secondes



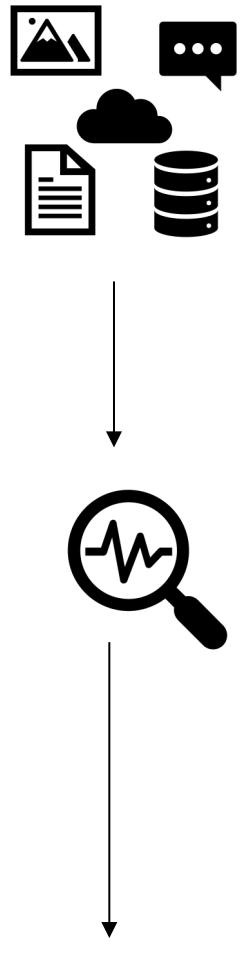
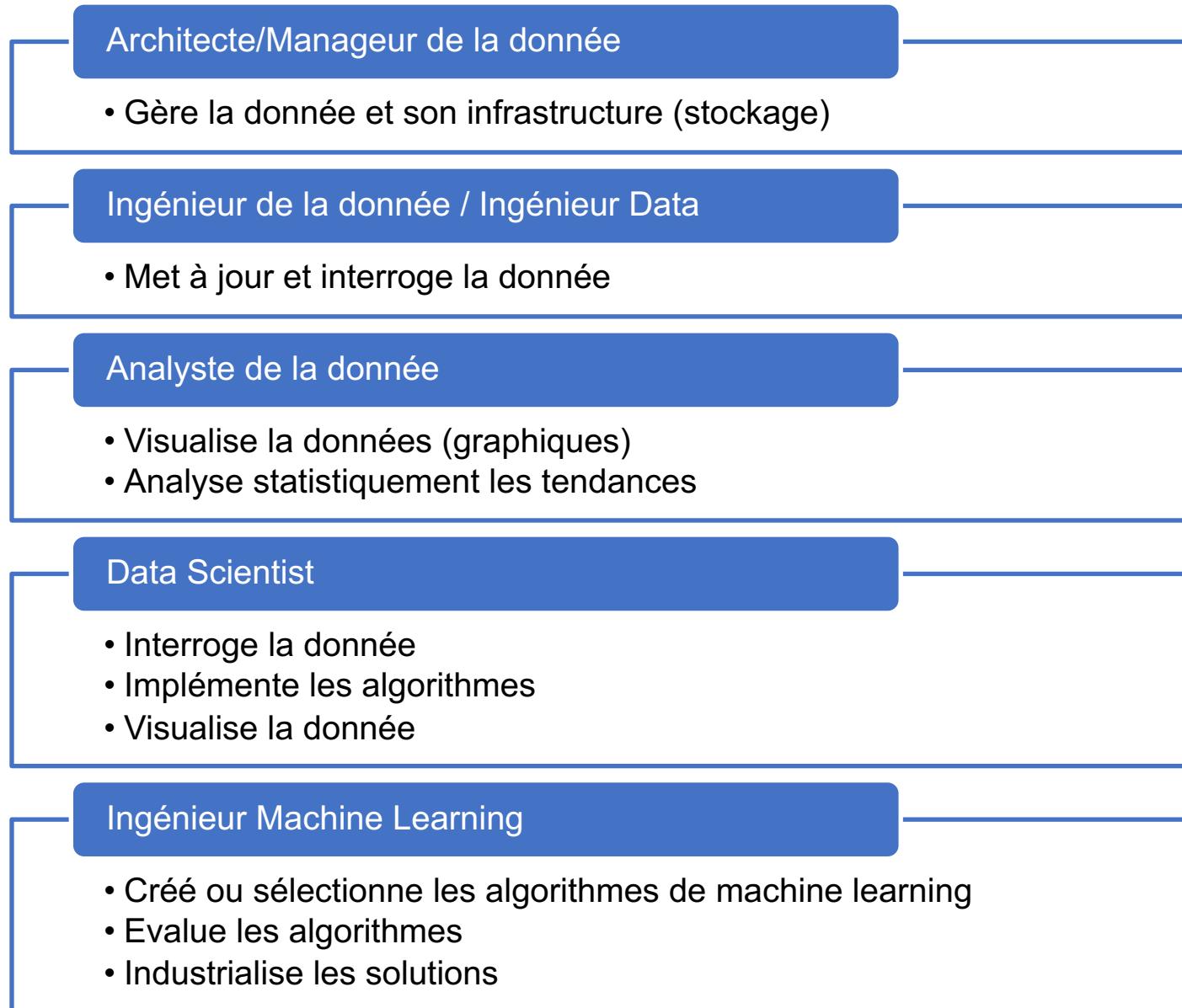
- Société Hugging Face

- Crée par trois français (Clément Delangue, Julien Chaumond et Thomas Wolf)
 - Plateforme pour le *Machine Learning* (apprentissage automatique) qui héberge modèles, jeu de données et applications web
 - Projet BigScience : modèle de langue appelé BLOOM (176 milliards de paramètres)



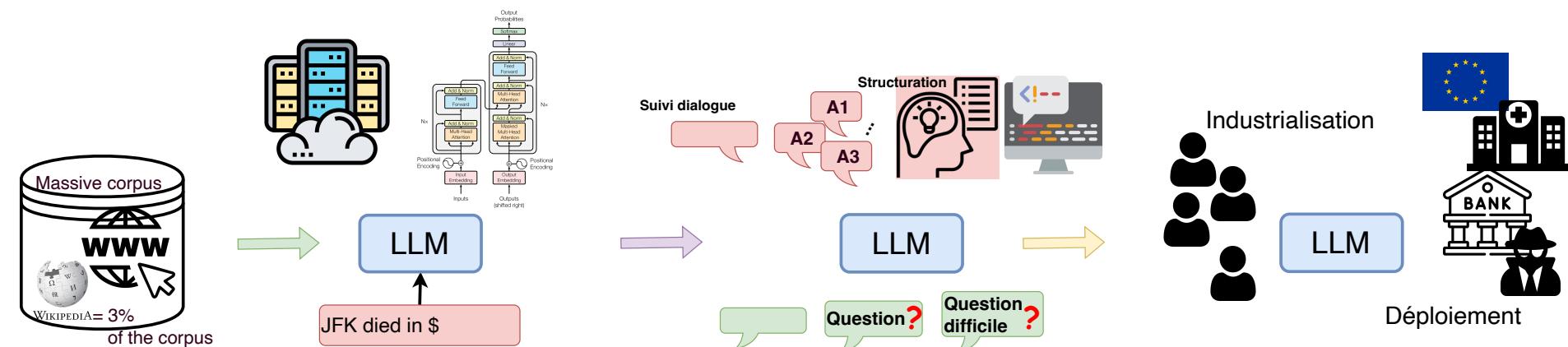
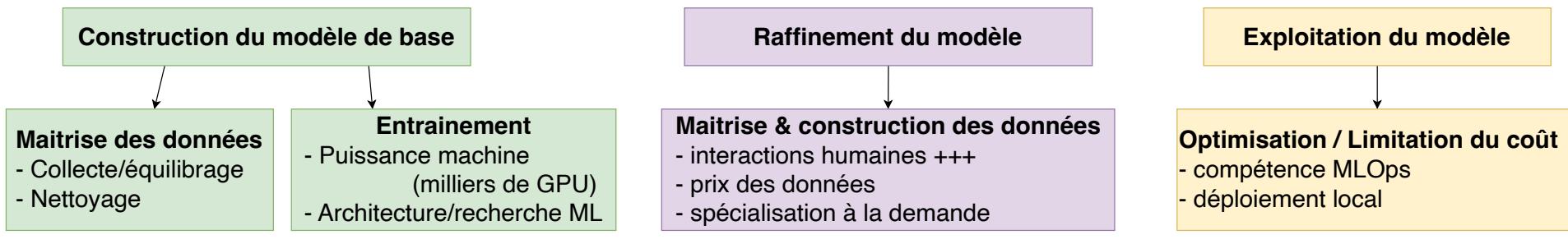
[1] CamemBERT: a Tasty French Language Model, Martin et al., 2020

[2] FlauBERT: Unsupervised Language Model Pre-training for French, Le et al, 2020



- Maitrise des données (massives) + données de raffinement
- Développement des modèles + spécialisation
- Construction des infrastructures

Chaine complète des modèles de langue



→ IA de confiance :

- Action humaine et surveillance/contrôle
- Robustesse et sécurité technique
- Protection de la vie privée et gouvernance des données
- Transparence
- Diversité, non-discrimination et équité
- Bien-être environnement et sociétal
- Responsabilité

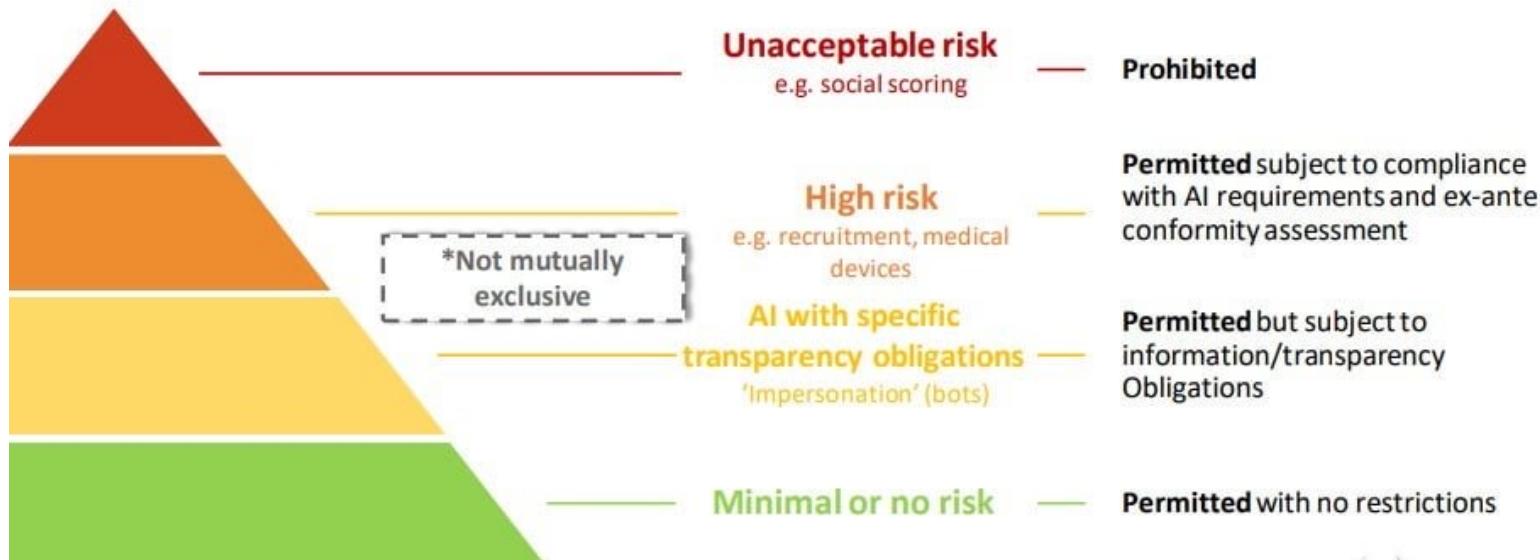


→ De nombreuses initiatives

→ Nationale : CERNA, CNPEN, 3IA

→ International : UNESCO, Déclaration de Montréal, OECD

A risk-based approach to regulation



Questions ?
